NASA Conference Publication 3130

Space and Earth Science Data Compression Workshop

Proceedings of a workshop held at the Snowbird Conference Center in Snowbird, Utah April 11, 1991

NASA

NASA Conference Publication 3130

Space and Earth Science Data Compression Workshop

James C. Tilton, *Editor*
NASA Goddard Space Flight Center
Greenbelt, Maryland

Proceedings of a workshop sponsored by the
National Aeronautics and Space Administration
and the IEEE Computer Society Technical Committee
on Computer Communications and held at
the Snowbird Conference Center in
Snowbird, Utah
April 11, 1991

NASA
National Aeronautics and
Space Administration
Office of Management
Scientific and Technical
Information Program
1991

FOREWORD

A "Space and Earth Science Data Compression Workshop" was held on April 11, 1991 in Snowbird, Utah. This document is the proceedings from the workshop. This workshop was held in conjunction with the 1991 Data Compression Conference (DCC'91), which was held at the same location April 8-10, 1991. Both DCC'91 and this workshop were a follow-up on a "Scientific Data Compression Workshop" that was held May 3-5, 1988 in Snowbird, Utah. The proceedings of the 1988 workshop can be obtained by contacting James C. Tilton (for address, see Appendix), as can additional copies of these proceedings.

This workshop explored the opportunities for data compression to enhance the collection and analysis of space and Earth science data. In seeking to identify the most appropriate data compression approaches, the workshop focused on the scientists' data requirements, as well as constraints imposed by the data collection, transmission, distribution and archival systems.

The workshop consisted of several invited papers, followed by group discussions. Two invited papers described information systems for space and Earth science data. These papers addressed present and proposed configurations, focusing on the constraints imposed by collecting, transmitting, distributing and archiving the data. One paper focused on the Earth Observing System Data and Information System (EOSDIS), and the other on the data system for the CRAF-Cassini Project.

Four invited papers depicted analysis scenarios for extracting information of scientific interest from data collected by Earth-orbiting and deep-space platforms. Examples discussed included data expected from the Moderate Resolution Imaging Spectrometer (MODIS), Synthetic Aperture Radar (SAR) data, observation data from spacecraft investigating space plasma physics, and data from microgravity experiments inboard the space shuttle or proposed space station.

A final invited paper was a general tutorial on image data compression.

After the invited papers, most workshop participants joined one of three discussion groups, namely:

(i) Data Compression for Data Archival and Browse/Quick Look,
(ii) Data Compression for Near Earth and Deep Space to Earth Transmission, and
(iii) Techniques for Containing Error Propagation in Compression/Decompression Schemes.

The first goal of each discussion group was to examine the potential for data compression to address data storage and transmission constraints found throughout the domain of NASA missions. The second goal was to recommend specific actions directed at enabling mission use of appropriate data compression technologies to overcome these constraints. These recommendations are summarized in the following section.

Discussion Group Recommendations

Users and developers of data compression technologies should be brought in closer communication within NASA, and with academia, industry and other government agencies. A data compression working group, newsletter, and/or electronic bulletin board should be established.

NASA should provide test data sets and examples of analysis scenarios to the data compression research community. These data sets should cover a broad range of NASA applications, concentrating on high data volume cases, and cases requiring high transmission bandwidth between the sensors and Earth, and across communications networks on Earth.

NASA should use lossless data compression wherever possible to improve communications and storage capacity. NASA should continue working with the Consultative Committee for Space Data Systems to define lossless data compression standards, so that space qualified hardware can make maximum use of commonality.

NASA should encourage the application of data compression techniques to data browse and archival. Research is needed especially in developing "smart" browse techniques, in which the lossily compressed data retains most of the essential scientific information for a rough, but informative, scientific analysis of the data. Key to this research is the participation of Earth and space scientists who would evaluate the decrease of science value due to the distortions introduced by lossy compression, and the increase in science value due to increased temporal, spectral and measurement resolution increased coverage. Other research is required into techniques for integrating the "smart" browse data into the data archive access system.

NASA should develop and select approaches to high-ratio compression of operational data such as voice and video.

NASA should examine the use of lossy compression techniques in combination with A-D conversion. The current approach using a uniform (or perhaps companded) quantizer followed by lossless compression (if compression is employed) is suboptimal. An example of employing lossy compression techniques to optimize this process would be convert the analog signal into vector codes, such as done in vector quantization (a form of lossy compression). Vector quantization design techniques could then be employed to tailor the overall source code to characteristic of the data being encoded.

NASA should examine new data compression approaches, such as combining source and channel encoding, where high-payoff gaps are identified in currently available schemes.

NASA should pursue research into the optimal integration of error containment and error correction with data compression. Here, the data compression scheme aids in error detection and subsequent error correction.

NASA should develop data compression integrated circuits for a few key approaches identified in the preceding recommendations.

Finally, we recommend that NASA should make the pursuit of research in these and other promising areas related to the compression of space and Earth science data an area of emphasis in one or more future solicitations (e.g., NASA Research Announcement) under the Applied Information Systems Research Program and/or other appropriate NASA program.

Acknowledgment

The organization of this workshop was supported by the Office of Aeronautics, Exploration and Technology, NASA Headquarters, Washington, DC.

Workshop Organizers

James C. Tilton and
Daniel E. Erickson

SPACE AND EARTH SCIENCE DATA COMPRESSION WORKSHOP

April 11, 1991 - Snowbird, Utah

8:00am: **Welcome and opening remarks** from Workshop organizers: Dr. James C. Tilton of NASA GSFC and Dr. Daniel E. Erickson of NASA JPL.

8:10am: **NASA Headquarters welcome:** Mr. Joseph Bredekamp, Office of Space Science and Applications, NASA Headquarters.

Morning session I - Science Data Systems: 8:15 - 9:45am

8:15am: **Overview of the EOS Data and Information System**, Dr. Jeff Dozier, Earth Observing System (EOS) Project Scientist, NASA Goddard Space Flight Center, Greenbelt, MD.

9:00am: **Data Compression - The End-to-End Information System Perspective for NASA Space Science Missions**, Mr. Wallace Tai, End-to-End Information System Engineer, CRAF-Cassini Project, Jet Propulsion Laboratory, Pasadena, CA.

Morning session II - Science Data Requirements: 9:45am - NOON

9:45am: **The Moderate Resolution Imaging Spectrometer: An EOS Facility Instrument Candidate for Application of Data Compression Methods**, Dr. Vincent Salomonson, Team Leader for the Moderate-Resolution Imaging Spectrometer, NASA Goddard Space Flight Center, Greenbelt, MD.

10:15am: **Break**

10:30am: **SAR Data Compression: Applications, Requirements and Designs**, Dr. John C. Curlander, Jet Propulsion Laboratory, Pasadena, CA.

11:00am: **Scientific Requirements for Space Science Data Systems**, Dr. Ray Walker, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA.

11:30am: **Microgravity Science Requirements and the Need for Data Compression**, Mr. William Hartz, Principal Engineer for diagnostics systems of the Combustion Experiments Module at the NASA Lewis Research Center, Analex Corporation, Cleveland, OH.

Lunch Session - Data Compression Approaches: NOON - 1:15pm

NOON: **Break**

12:30pm: **Image Compression**, Dr. Robert Gray, Professor, Electrical Engineering Department, Stanford University.

Afternoon session I - Group Discussions: 1:15 - 4:00pm

Afternoon session II - Summary Group Reports: 4:00 - 4:30pm

CONTENTS

INVITED PAPERS

DISCUSSION GROUP REPORTS

APPENDIX

omit to
p5

INVITED PAPERS

THE INVITED PAPERS

The Morning and Lunch Sessions of the Space and Earth Science Data Compression Workshop consisted of the presentations of seven invited papers on Science Data Systems, Science Data Requirements, and Data Compression Approaches. Papers based of five of those seven presentations follow. Abstracts for the two papers not included here can be found in the Proceedings of the Data Compression Conference, as given in reference form below:

[1] Jeff Dozier, "Overview of the EOS Data and Information System," *Proceedings of the Data Compression Conference*, Snowbird, Utah, April 8-11, 1991, p. 472.

[2] Robert M. Gray, "Image Compression," *Proceedings of the Data Compression Conference*, Snowbird, Utah, April 8-11, 1991, pp. 474-5.

PRECEDING PAGE BLANK NOT FILMED

N92
12426
UNCLAS

DATA COMPRESSION - THE END-TO-END INFORMATION SYSTEMS PERSPECTIVE FOR NASA SPACE SCIENCE MISSIONS

Wallace Tai
Jet Propulsion Laboratory
Pasadena, CA 91109

Abstract. The unique characteristics of compressed data have important implications to the design of space science data systems, science applications, and data compression techniques. The sequential nature or data dependence between each of the sample values within a block of compressed data introduces an error multiplication/propagation factor which compounds the effects of communication errors. The data communication characteristics of the on-board data acquisition, storage and telecommunication channels may influence the size of the compressed blocks and the frequency of included re-initialization points. The organization (i.e. size and structure) of the compressed data are continually changing depending on the entropy of the input data. This also results in a variable output rate from the instrument which may require buffering to interface with the spacecraft data system. On the ground, there exist key trade-off issues associated with the distribution and management of the science data products when data compression techniques are applied in order to alleviate the constraints imposed by ground communication bandwidth and data storage capacity.

Missions that anticipate utilizing data compression could improve their information throughput efficiency by influencing sensor and instrument design that are synergistic with the spacecraft data acquisition and data management schemes, the science application requirements (including quick look data analysis), and characteristics of the data collection and downlink communication channels. In summary, data compression, its application and effects, must be understood in the context of an end-to-end information system.

1. Introduction

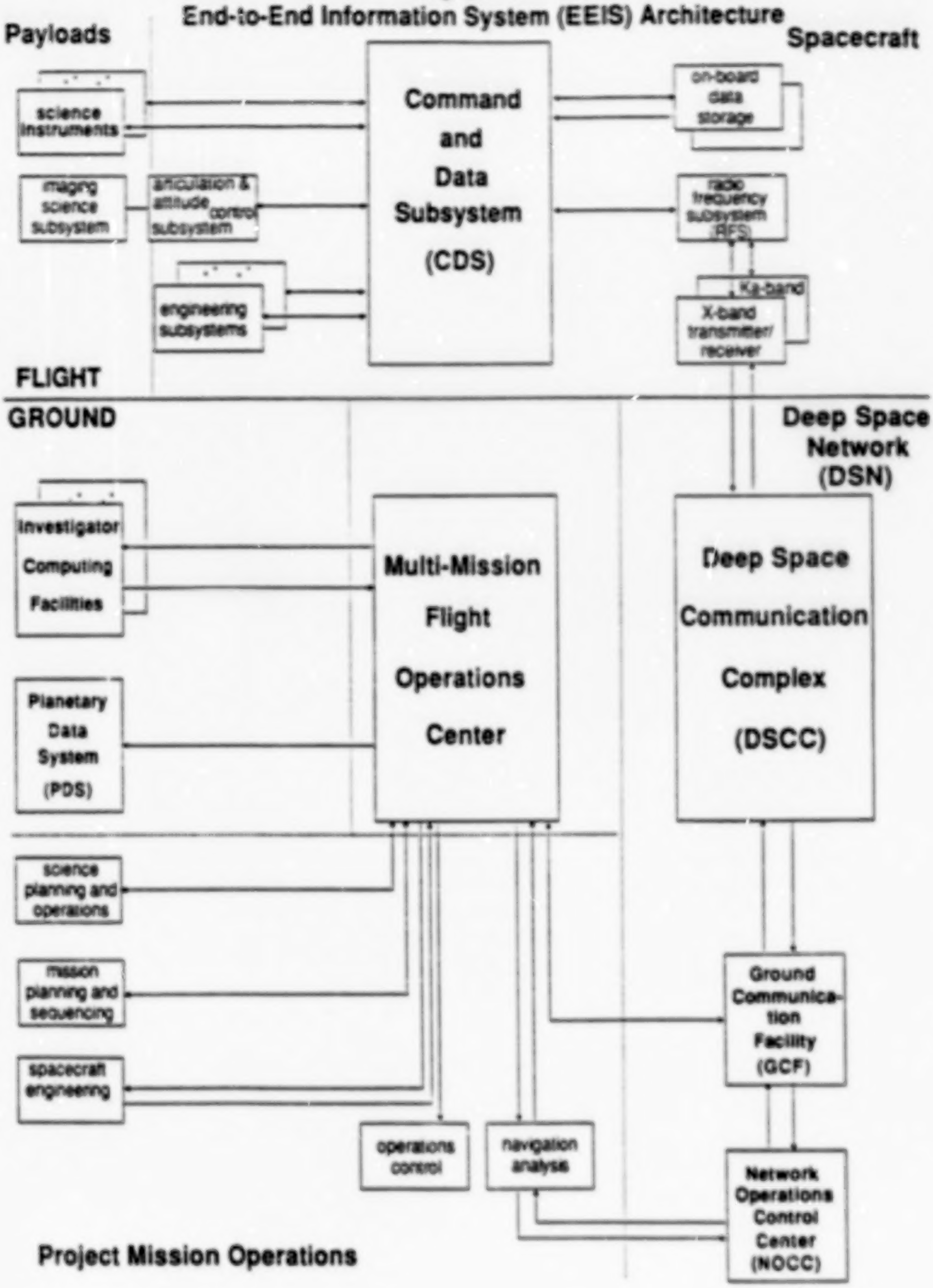
This paper gives an overview on the architecture of the end-to-end information system (EEIS) for NASA planetary missions, its major constraints, and the effects on the system due to the application of data compression techniques. Issues surrounding data compression cannot be viewed as technological issues alone, nor can they be confined to the elements where compression and/or decompression take place. For a NASA planetary mission, data compression has profound implications to science, mission design, flight and ground data systems, and mission operations. It is believed that the application of data compression as a technology onto space missions environment must take into account its propagating effects on elements throughout the EEIS. As such, a system engineering perspective is crucial to the successful implementation of a system architecture using data compression.

2. Architectural Overview of the End-to-End Space Science Data System

The End-to-End Information System (EEIS) for a NASA planetary mission can be viewed as a set of functions, distributed throughout the flight and ground systems that operate cooperatively to collect, transport, process, store, and analyze the data and information in the mission. Functionally, the EEIS can be decomposed into two processes: a downlink process and a uplink process. Architecturally, the EEIS consists of the following key physical components (Refer to Figure 1):

Figure 1

End-to-End Information System (EEIS) Architecture



Flight Elements -

Science instruments
Command and Data Subsystem (CDS) including the on-board mass storage
Radio Frequency Subsystem (RFS) including its transmitters

Ground Elements -

Deep Space Network (DSN)
The Multi-Mission Flight Operations Center at JPL
Project-specific Mission Operations Elements:
-- Science planning and operations
-- Mission planning
-- Scheduling and sequencing
-- Spacecraft engineering
-- Operations control
-- Navigation analysis
Planetary Data System (PDS)

2.1 Downlink Description

The downlink process begins at each science instrument or spacecraft engineering subsystem acquiring science data and/or engineering data. The various instruments and subsystems will concurrently output its data in the form of CCSDS source packets to the CDS. All data packets will be assembled by the CDS into CCSDS transfer frames for storing on-board before the transmission via the downlink channel provided by the Deep Space Network (DSN).

On the ground, the DSN Ground Communication Facility (GCF) is responsible for delivering the received data at the tracking stations, i.e. the Deep Space Communication Complexes (DSCC), to the Multi-Mission Flight Operations Center and DSN Network Operations Control Center (NOCC) at JPL. At the Multi-Mission Flight Operations Center, spacecraft engineering data and instrument engineering data are processed for spacecraft and instrument health monitoring. Furthermore, science data, in particular the imaging data, will be processed for science analysis in support of science and mission operations. The DSN NOCC will perform radiometric data conditioning, VLBI correlation, and generate earth rotation calibration information. It also has the responsibility for monitoring and assessing the performance of each DSCC. The facilities, tools, and data provided the Multi-Mission Flight Operations Center and NOCC will be used by the flight project-specific mission operations elements such as science planning and operations, mission planning, spacecraft engineering, navigation analysis, and operations control to perform downlink-related analysis functions. During the operational phase, the various science teams will access the science data and ancillary data to perform science analysis from their home facilities. The science data, ancillary data, and associated engineering data generated and maintained by the Multi-Mission Flight Operations Center will eventually be transferred to and archived at the Planetary Data System (PDS) for access by the general planetary science community.

2.2 Uplink Description

The uplink process begins with the development of a long term mission operations plan and a science planning guide for each mission phase or key subphase based on the mission plan. These long term plans will then be used to generate a set of short term plans such as a navigation plan, a conflict-free science plan, and an integrated mission timeline. All these planning activities are performed by the project-specific mission operations elements, i.e. science planning and operations, mission planning, spacecraft engineering, and navigation analysis, in a coordinated fashion using the data system services

provided by the JPL Multi-Mission Flight Operations Center. The sequences for a mission phase or subphase will then be developed. The end result of the sequence generation activity is the weekly command load ready for uplink. To deliver the command load to the spacecraft from JPL via DSN GCF and DSCC, a set of CCSDS telecommand operations procedures will be executed at both the Multi-Mission Flight Operations Center and the CDS to ensure successful delivery and accountability.

On the spacecraft, the CDS will manage the execution of sequences by the spacecraft subsystems and deliver the commands to the instruments for execution. As part of the uplink process, the CDS is also responsible for on-board control of all flight elements in response to certain natural events and fault protection in response to significant anomaly conditions.

3. General Constraints of an End-to-End Information System

In general, the data and information system for a planetary mission is more constrained than its counterpart for an earth mission. In particular, on the flight side, the primary constraints such as power, mass, thermal control, and positioning for the planetary missions have direct effect on the design of the data systems on the spacecraft and instruments. To the science data collection process, the results is limitations on the data rate, processing power, physical memory size, on-board data storage capacity, and local communication bandwidth.

The quantity and quality of data transmitted over the space-to-ground communication channel are limited by the telecommunication link performance. For planetary missions, the distance between the spacecraft and the earth as well as transmitter power, weather conditions, background noise from target body, and other factors is a very important parameter for the determination of allowable data rates and error rates. In addition, from the mission operations perspective, the availability of receiver stations on the ground, in terms of their tracking time and relative geometry to the spacecraft, is also a constraint considering the fact that the ground stations of Deep Space Network (DSN) as a multi-mission resource have always been over-subscribed.

On the ground, as the computer technology advances there has been significant increase in demands on the ground processing and archiving systems. Pertinent to planetary missions, two chief demand-driven constraints are observed:

(1) The timely delivery of science data products in large volumes to a community of geographically distributed investigators is still considered a difficult task due to the limited bandwidth of the ground communication networks.

(2) As more remote sensing data become available to the general planetary science community, the rapid access to science data products in the data archive system is constrained by the need to have prior knowledge about the data formats and information contents (i.e. both in syntax and semantics) about the products.

Data compression as a technology has long been employed in the planetary missions as of the solutions to these constraints in order to maximize science return. In the course of its application, there has been precious lessons learned. The following sections summarize some of our engineering experience in this area.

4. Effects of Data Compression on End-to-End Information System

In general, compressed data has the following characteristics:

(1) Reduced data volume
(2) Asynchronous output data
(3) Variable length data
(4) Increased sensitivity to noise (or transmission errors)

These characteristics have important engineering implications to the EEIS. There are benefits and added complexity to the EEIS. Clearly, benefits gained by the EEIS through the use of data compression are primarily due to general characteristics (1):

Reduce the overall buffer size requirements throughout the breadth of the EEIS

For planetary missions, this is particularly true for those high-rate instruments employing data compression techniques to acquire observation data. Not only the instrument internal buffer size for science read-out data but also the overall telemetry collection buffer size on the spacecraft is reduced. As mentioned in Section 3, since the on-board memory size for planetary missions has always been a constrained resource reducing buffer size through sensor data compression certainly offers a viable approach to getting around this constraint.

Reduce communication data rate requirements

From the data transport perspective, the compressed data also reduces the communication data rate requirements by providing higher entropy in the data. For planetary missions, the beneficiaries are primarily the communication line between the instruments and spacecraft, the space-to-ground link, and ground communication network which carries the data to the ground system where decompression of the data is performed.

In data archive system environment, data in compressed form have been used for product distribution to minimize the medium capacity requirement. The application of compression techniques to generate browse data sets for near real-time distribution to the users has also helped the data archive system to overcome the constraint imposed by the need to have prior knowledge about the data formats and information contents about the products.

Increase coverage and/or resolution of the instruments

To the science investigators, data compression offers the flexibility for the instrument to compact the sensor data by reducing the number of bits required for each sample so that a larger area of coverage can be achieved by the instruments.

Accommodate the tailoring of data products generated for a specific application (through the use of lossy compression).

On the other hand, general characteristics (2), (3), and (4) inevitably add certain complexities to the system:

More stringent communications quality and continuity requirements for transported data

There is a sequential relationship between each of the sample values within a block of compressed data output. This relationship introduces an error multiplication/propagation factor which compounds the effects of communications errors. The error introduced by the communication channel in a sample value may invalidate all the subsequent sample values in the

same block. Consequently, more stringent data quality requirements must be levied on the EEIS. For planetary missions, typically the end-to-end bit error rate requirement on compressed data is 1×10^{-6} whereas for uncompressed data (in particular for certain circumstances where the data by their nature possess redundancy) it is 1×10^{-3}.

Added complexity in on-board buffer management due to variations in data compression profiles

As stated in Section 3, local communication bandwidth is also a constraint for spacecraft in planetary mission. The conventional telemetry data collection scheme on the spacecraft for planetary missions can be characterized by a deterministic approach where packets of data generated by the various instruments and spacecraft subsystems are picked up by the spacecraft in a time synchronized manner based on a priori knowledge about the outputs from these instruments and subsystems. This deterministic approach appears to be simple but is problematic to science instruments using data compression for maximizing their data returns. It requires the output data from each source remain constant during a telemetry collection "mode" based on the pre-defined parameters such as data rates, destinations, and packet lengths associated with all the instruments and subsystems during a period of interest. Compressed data, which is non-deterministic and variable in output rates, asynchronous in output timing, and variable in length, certainly does not land itself very well in this conventional telemetry collection environment. To compensate for this problem and make instruments compatible with the deterministic scheme, one of the methods is to include in each instrument a buffer management capability which allows it to match the variable data rates of the output from the compressor to the fixed data collection rate imposed by the spacecraft. However, there are two potential drawbacks in this remedy:

(1) When the output data rate during a pick-up cycle is lower than the scheduled and allocated data rate, filler data must be generated, negating some of the advantages of using data compression mentioned above.

(2) When the output data rate during a pick-up cycle is higher than the scheduled and allocated constant data rate, portion of the data in the instrument buffer will not be picked up by the spacecraft in time for the current cycle. The delayed transfer of bursty data, if persists through subsequent pick-up cycles, may eventually result in buffer overflow and data loss. To control the data loss, one may apply a lossy compression scheme as an option to force the compression ratios to a limit. An alternative is for the instrument to provide buffering capability accommodating long-term averaging of the data rates.

Obviously, in the context of the conventional telemetry data collection scheme, an important instrument design issue is the determination of the optimum fixed data rate as part of scheduling and allocation of on-board resources. It involves the trade-off between acceptable data loss and benefits gained for using compression but reduced because of filler data, and data rate allocation. For example, in order to avoid losing data, the fixed-rate scheme would have to allocate the maximum possible rate, negating the advantage of reducing communication data rate requirements offered by general characteristics (1).

A conclusion one may draw here is that even with a deterministic scheme it is difficult to expect a deterministic knowledge in the completeness of data collection. Under the resource constraints, the data loss will occur and there is no way to predict the amount of data loss.

An alternative to the conventional telemetry data collection scheme would be the data-driven approach which allows each instrument to output its data in variable length at variable rate asynchronously. In this non-deterministic scheme, at least three services must be provided by the spacecraft:

(1) The flight data system of the spacecraft will provide the rate buffering capability.

(2) Given a pool of consumable resources with certain margins allocated to each instrument, the flight data system must be capable of keeping track of the utilization of data rates and other related on-board resources, e.g. memory buffer, by all the instruments.

(3) The flight data system must be sufficiently robust to detect and respond to the "overdrawal" of data rate resource by any instrument.

On the instrument side, data rate is allocated to each instrument not in terms of a fixed, absolute number but in a range which may or may not vary as a function of time. The instrument must be capable of ensuring that its output data rate never exceed the upper bound of the range.

Overhead in uplink sequence development

During mission operations, a challenge encountered in developing the sequences for science data collection is the determination of the output data rate from the instruments using data compression. Assumption has to be made about the average compression ratio. In the case of the data-driven data collection approach, the stochastic property of resource utilization by each instrument and the more dynamic allocation of the collective, pooled resources must be modeled. Both average and worst-case situations will have to be evaluated. The amount of potential data loss as an additional parameter of the model also imposes extra complexity to the ground operations. To the science investigators, more options are available for them to make trade-off between the observation cost and science benefit by considering the competing factors such as data rates, data volumes, and data coverage. On the whole, the overhead in sequence development is caused by the added flexibility in flight offered by the more adaptive data collection design.

Increased computation required on-board and in ground systems

The process of compressing and decompressing data demands additional computations in both the flight and ground systems. Compressor performance must be compatible the readout rate of the sensors. Associated with this are a couple of key design issues:

(1) Location of the compressor - Should the spacecraft provide the compression (especially noiseless compression) as a service to all the instruments requiring compression on their data, or should each instrument contain a compressor as an integrated part of the instrument?

(2) Flexibility of the compressor - Can a flexible, generalized noiseless compressor be designed such that an off-the-shelf product can be available to reduce the development cost of the compressor chips?

5. Conclusions

For the future NASA planetary missions, more extensive application of data compression to the data and information systems seem to be dependent on the resolution of some of the system issues discussed above. There seems to be the need to carry out the following suggestions :

(1) Implementation of a data-driven telemetry collection scheme on the flight data systems.
(2) Extensive use of solid state recorder as rate buffering between the following processes on board:

On-board data collection
On-board data storage
Downlink

(3) Adopt a standard compressor for all NASA flight instruments requiring data compression service.

(4) Use of Reed-Solomon encoding on the downlink channel to minimize the effect of noise on the quality of compressed data.

N92

12427

UNCLAS

N92-12427

S3-35
51864

THE MODERATE RESOLUTION IMAGING SPECTROMETER: AN EOS FACILITY INSTRUMENT CANDIDATE FOR APPLICATION OF DATA COMPRESSION METHODS

P. 11

Vincent V. Salomonson
Earth Sciences Directorate
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

Abstract. The Moderate Resolution Imaging Spectrometer (MODIS) observing facility will operate on the Earth Observing System (EOS) in the late 1990's. It estimated that this observing facility will produce over 200 gigabytes of data per day requiring a storage capability of just over 300 gigabytes per day. Archiving, browsing, and distributing the data associated with MODIS represents a rich opportunity for testing and applying both lossless and lossy data compression methods.

1. Introduction

MODIS is a multispectral imaging system to be flown on the EOS in the late 1990's. The capability of the MODIS instrument derives from and expands upon some instruments that have been successfully flown on spacecraft or aircraft and used for many years to observe properties of the earth-atmosphere system and to develop data bases for studies of global change. These instruments are the Advanced Very High Resolution Radiometer (AVHRR) and the High Resolution Infrared Sounder (HIRS) being currently flown on the NOAA operational meteorological satellites, the Coastal Zone Color Scanner (CZCS) flown on the Nimbus-7, the Landsat Thematic Mapper, and various aircraft scanners.

The MODIS system will be composed of two cross-track scanning instruments. One instrument is called MODIS-N (nadir), indicating a multispectral scanner that will not be tilted and provide a continuous cross-track scan. The other instrument is called MODIS-T (tilt), indicating a scanner that will allow the cross-track scan to be tilted 50° fore and aft. The MODIS-N will have 36 spectral bands covering spectral bands in the visible, near-infrared (0.7-1 microns), the short-wave infrared (1-3 microns), and the thermal infrared (3-15 microns). Tables 1 and 2 summarize MODIS-N and Table 3 summarizes MODIS-T. The purposes of the bands in Table 2 are only indicative and not complete. Further details concerning MODIS-T and MODIS-N instrumentation are given in Salomonson [2] and Magner and Salomonson [1].

Data volume coming from the MODIS observing facility will be on the order of a terabit of data per day depending upon the total time the instruments are on and the amount of ancillary information acquired concerning spacecraft attitude, etc. It is quite appropriate, therefore, to consider the MODIS as a case study for data compression methods that would reduce the volume of data involved for archiving, distribution, and analysis. This paper will describe in more quantitative detail the volumes and rates of data associated with the MODIS. The objective of this discussion is to provide those interested in applying data compression methods to MODIS, as a relevant and challenging example, with particulars that should help in assessing the magnitude and complexities of the task.

2. MODIS Data Volumes and Rates

In the broadest sense, it is envisioned that data compression methods might be most appropriately applied to MODIS data for the purposes of data storage, data distribution, providing a browsing capability, or facilitating the direct broadcast of MODIS data to terminals on the ground where only a subset or specific parameter of the data is needed or where reduced data volume or rate is needed in order to be compatible with limited receiving or processing capabilities. It is assumed in this paper that all information for MODIS must be retained in processing, at least through level 1, if not to level 2. The

reason for being resolute with regard to the level 1 data processing is that from these data are derived all level 2 products. The level 2 products, of which there will be as many as 100, must utilize all the radiometric and calibration fidelity in level 1. Even in the storage of MODIS data, particularly in the case of level 1, all information must be retained. This, therefore, indicates that only lossless data compression methods should be applied to data storage for level 1, and perhaps level 2 and above. Lossy data compression methods are deemed, at this point in the author's understanding, appropriate for producing browse data or for very specific applications wherein the loss of information can be tolerated.

Figure 1 depicts the overall data flow for the MODIS. This figure shows that MODIS data will flow from the EOS platform through the Tracking and Data Relay Satellite (TDRS) to the Customer and Data Operations Systems (CDOS) and into the Goddard Space Flight Center (GSFC) Distributed Active Archive Center (DAAC). In the DAAC, data will be processed after algorithms have been developed and checked for accuracy and quality by the MODIS Science Team Members using the several and distributed Team Member Computing Facilities (TMCF's) and the MODIS Team Leader Computing Facility (TLCF). When products are produced in the appropriate DAAC, they will be archived and distributed to the scientific community and any part of the public at large that desires to use MODIS data. The distribution of the data will occur through the EOS Data and Information System (EOSDIS) electronic network that exists when MODIS becomes operational.

As further detail, Table 4 shows specifics concerning calculated data rates and volumes associated with MODIS-N and -T. In this table, it is worth noting that the 13th bit for MODIS-T data is only included to flag the gain used in sending MODIS-T data to the ground (see [1]). In going from level 1A (calibration and navigation information provided in the header, but not applied) to level 1B (calibration information data applied and pixel location available), the increased volume is due principally to converting the 12-bit information to 16 bits and adding navigation, calibration, and browse information.

Table 5 shows the archiving requirements in gigabytes per day for MODIS-N and -T. These are rough estimates based on preliminary estimates using existing or planned algorithms for the principal products to be derived from the MODIS. In many cases these estimates are based upon experience from the heritage instruments indicated in the Introduction. From a similar perspective, the expected lines of code estimates shown in Table 6 have been derived. In general, the bulk of the effort for producing lines of code and storing the results falls in producing at-satellite radiances (levels 1A and 1B) and in producing water leaving and land leaving radiances.

Table 7 shows the estimated load on the data distribution system in providing MODIS data to other DAAC's in the EOSDIS. The other main DAAC, besides the GSFC DAAC, is at the EROS Data Center (EDC) in Sioux Falls, South Dakota. At the EDC, all level 2 products produced over land areas will be archived and all level 3 land products will be produced and distributed from the EDC. Other key DAAC's are at the Langley Research Center in Virginia and the National Snow and Ice Data Center in Boulder, Colorado. The assumptions as to the fraction of MODIS data that will go to these DAAC's is provided in the third column of Table 7.

With regard to direct broadcast of MODIS data and the volume of MODIS browse data, the following statements are provided. For direct broadcast it has been assumed that 100 percent of the raw data would be involved, but, in this instance, on-board data compression techniques could be examined. If such an approach is to be used, however, that must be decided soon (i.e., in a year or 2) in order for it to be implemented with the instrument or on the EOS spacecraft. The situations surrounding direct broadcast from the EOS are relatively undefined, but one may assume that a 15 megabit direct broadcast link will be available to be shared among the instruments. MODIS, of course, has the potential for occupying a large share of this capability unless data compression is applied appropriately. In the case of browse data, by assuming that browse data will be comprised of 5 percent of levels 1B, 2, and 3,

END-TO-END MODIS DATA FLOW FROM INSTRUMENT TO USERS

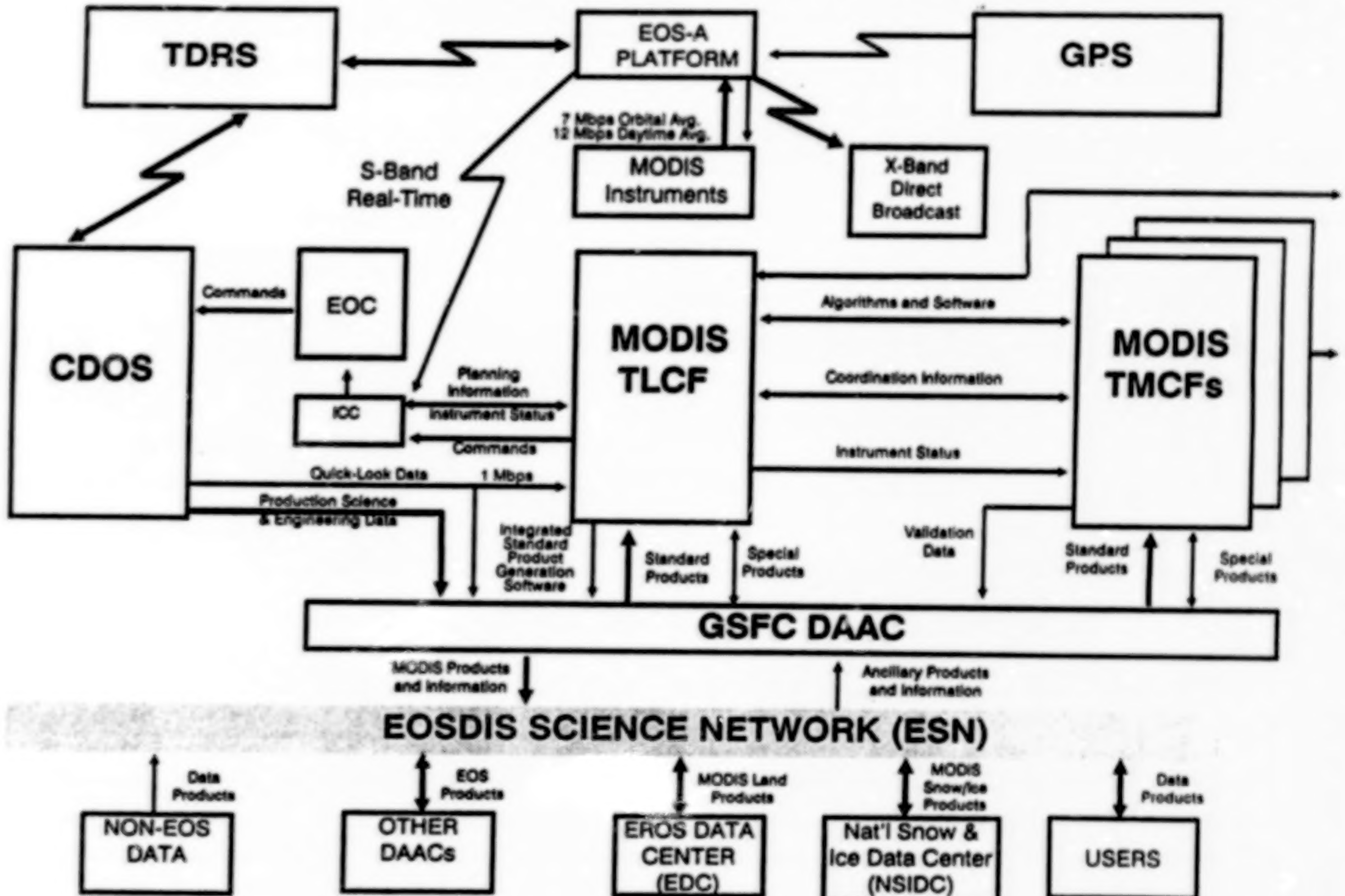


FIGURE 1

resulting in an estimate of about 12 gigabytes of data per day. Browse data is a prime candidate for applying lossy compression methods.

In the case of archiving MODIS data, it has already been indicated that no information should be lost in archiving the data. However, lossless compression methods could, and perhaps should, be applied that allow the progressive extraction of archived data at various levels of accuracy depending upon the amount of information actually needed. This means that if the data are compressed appropriately, one could access the archive and extract first-order information. If this initial extraction indicates further information is needed, another pass through the compressed archive could result in higher-order information. Ultimately it appears that data compression methods are available for archived data wherein the complete information available in the original data stream ultimately could be retrieved.

3. Summary and Conclusions

The MODIS provides a rich opportunity for applying data compression methods for archiving, browsing, and distribution. Lossless methods should be developed for archiving that allow eventual extraction of all the information contained in the MODIS. Lossy methods can very appropriately be applied in order to browse MODIS data and distribute it for quick-look analyses. The challenges include costs of developing and applying data compression methods including associated hardware costs, the availability of off-the-shelf versus special purpose hardware and software, demonstrating reliability and low risk of losing information for lossless methods plus, in many cases, making the application of data compression transparent to the average user.

Acknowledgments

The author wishes to acknowledge the assistance of Dr. A. Fleig, Dr. D. Han, and members of the Research and Data Systems Corporation (RDC) staff for developing the tables and figure in this paper. Kelly Pecnick prepared the final drafts of the paper. Her help and expertise are very much appreciated.

References

[1] Magner, T. J., and V. V. Salomonson, 1991: Moderate Resolution Imaging Spectrometer-Tilt (MODIS-T). *International Journal of Imaging Systems and Technology*, Vol. 3.

[2] Salomonson, V. V., 1991: Execution Phase (C/D) Spectral Band Characteristics of the EOS Moderate Resolution Imaging Spectrometer-Nadir (MODIS-N) Facility Instrument. *Advances in Space Research*, Vol. 3, pp. 231-236.

TABLE 1

MODIS-Nadir (N) Summary

PARAMETERS	DESIGN SPECIFICATIONS OR EXPECTED PERFORMANCE
PLATFORM ALTITUDE	705 KM
IFOV (no. of bands @ IFOV)	29 @ 1000 M 5 @ 500 M 2 @ 250 M
SWATH	110 DEG/2330 KM
SPECTRAL BANDS	36 BANDS TOTAL (19/0.4-3.0 μm; 17/3-15 μm)
RADIOMETRIC ACCURACY	5% ABSOLUTE, < 3 μm 1% ABSOLUTE, > 3 μm (possibly < 0.3%) 2% REFLECTANCE
QUANTIZATION	12 BIT
POLARIZATION SENSITIVITY	2% MAX, < 2.2 μm
MODULATION TRANSFER FUNCTION	0.3 AT NYQUIST
S/N PERFORMANCE (70 DEGREE SOLAR ZENITH/OCEANS)	830:1 (443 nm) 745:1 (520 nm) 503:1 (865 nm)
NEDT PERFORMANCE (THERMAL BANDS) @ 300 DEG K/WINDOW BANDS	LESS THAN 0.05
SCAN EFFICIENCY	(TO BE DETERMINED)
INTEGRATION TIME	(TO BE DETERMINED)
SIZE (APPROX)	1 X 1.6 X 1 M
WEIGHT	APPROX 200 kg
POWER	250 w
PEAK DATA RATE	11 MBS (daytime)
DUTY CYCLE	100%

TABLE 2

MODIS-N Bands

BAND	CENTER *	IFOV (m)	WIDTH	PURPOSE
	LAND AND CLOUD BOUNDARIES BANDS			
1	659	250	50	VEG CHLOROPHYLL ABS LAND COVER TRANS.
2	865	250	40	CLOUD AND VEGETATION LAND COVER TRANSF.
	LAND AND CLOUD PROPERTIES BANDS			
3	470	500	20	SOIL, VEG DIFFRNCS
4	555	500	20	GREEN VEGETATION
5	1240	500	20	LEAF/CANOPY PROPRTIES
6	1640	500	20	SNOW/CLOUD DIFFRNCES
7	2130	500	50	LAND & CLOUD PROPRTIES
	OCEAN COLOR BANDS			
8	415	1000	15	CHLOROPHYLL
9	443	1000	10	CHLOROPHYLL
10	490	1000	10	CHLOROPHYLL
11	531	1000	10	CHLOROPHYLL
12	565	1000	10	SEDIMENTS
13	653	1000	15	SEDIMENTS, ATMOSPHERE
14	681	1000	10	CHLOR. FLUORESCENCE
15	750	1000	10	AEROSOL PROPERTIES
16	865	1000	15	AEROSOL/ATM PRPRTS
	ATMOSPHERE/CLOUD BANDS			
17	905	1000	30	CLOUD/ATM PROPERTIES
18	936	1000	10	CLOUD/ATM PROPERTIES
19	940	1000	50	CLOUD/ATM PROPERTIES
	THERMAL BANDS			
20	3.75	1000	0.18	SEA SURFACE TEMP
21	3.75	1000	0.05	FOREST FIRES/VOLCANOES
22	3.96	1000	0.05	CLOUD/SFC TEMPERATURE
23	4.05	1000	0.05	CLOUD/SFC TEMPERATURE
24	4.47	1000	0.05	TROP TEMP/CLD FRACTION
25	4.52	1000	0.05	TROP TEMP/CLD FRACTION
26	4.57	1000	0.05	TROP TEMP/CLD FRACTION
27	6.72	1000	0.36	MID-TROP HUMIDITY
28	7.33	1000	0.30	UPPER-TROP HUMIDITY
29	8.55	1000	0.30	SFC TEMPERATURE
30	9.73	1000	0.30	TOTAL OZONE
31	11.03	1000	0.50	CLOUD/SFC TEMPERATURE
32	12.02	1000	0.50	CLOUD/SFC TEMPERATURE
33	13.34	1000	0.30	CLD HEIGHT & FRACTION
34	13.64	1000	0.30	CLD HEIGHT & FRACTION
35	13.94	1000	0.30	CLD HEIGHT & FRACTION
36	14.24	1000	0.30	CLD HEIGHT & FRACTION

* BAND CENTER AND BANDWIDTH ARE IN NANOMETERS FOR BANDS 1-19 AND MICROMETERS FOR BANDS 20-36

TABLE 3

MODIS-Tilt (T) Summary

PARAMETERS	DESIGN SPECIFICATIONS OR EXPECTED PERFORMANCE
PLATFORM ALTITUDE	705 KM
IFOV	1.4 MRAD (1.1 KM)
SWATH	90 DEG/1500 KM
SPECTRAL BANDS (10-15 nm WIDTH)	32 (400-880 nm.) (AREA ARRAY)
DYNAMIC RANGE	Lmax 95% @ 22.5 deg solar zenith angle
RADIOMETRIC ACCURACY	5% absolute 2% relative to the sun
QUANTIZATION	12 BIT
POLARIZATION SENSITIVITY	<2.3 % (< 20 deg tilt)
MODULATION TRANSFER FUNCTION	0.3 AT NYQUIST
S/N PERFORMANCE (SPEC) (70 DEGREE SOLAR ZENITH)	835:1(440 nm) 685:1(625 nm) 400:1(845 nm)
NEDT PERFORMANCE (THERMAL BANDS)	N/A
SCAN EFFICIENCY	25 %
INTEGRATION TIME	1.127 MSEC (COMPOSITE MODE)
COLLECTING APERTURE (DIA)	34 MM
SIZE (APPROX)	75 X 140 X 100 cm
WEIGHT	~170 kg
POWER	~130 w
PEAK DATA RATE	~3 mbps (day)
DUTY CYCLE	DAYTIME/100%

TABLE 4

MODIS-N and MODIS-T Data Rate and Volume Estimates

Earth Radius (km)	6371
Satellite Altitude (km)	705
Orbital Period (min)	98.9
Modis-N # 1000 m REF channels	12
Modis-N # 500 m REF channels	3
Modis-N # 250 m REF channels	2
Modis-N # 1000 m TIR channels	17
Modis-N # 500 m NIR channels (1.6, 2.1 nm)	2
Modis-T # 1.1 km REF channels	32
MODIS-N # bits/REF channel	12
MODIS-N # bits/TIR channel	12
MODIS-T # bits/REF channel	13
MODIS-N REF Duty Cycle	50%
MODIS-N TIR Duty Cycle	100%
MODIS-T REF Duty Cycle	45%
MODIS-N # Along-track IFOVs	8
MODIS-T # Along-track IFOVs	30
MODIS-N # Detectors	648
MODIS-T # Along-track detectors	30
MODIS-N # Maximum scan angle (deg)	55
MODIS-T # Maximum scan angle (deg)	45
MODIS-N # IFOV FWHM (deg)	8.13E-02
MODIS-T # IFOV FWHM (deg)	8.94E-02
MODIS-N # pixels along-scan/on-Earth	1354
MODIS-T # pixels along-scan/on-Earth	1007
MODIS-N Scan Period (sec)	1.2
MODIS-T Scan Period (sec)	4.6
MODIS-N VIS Data (megabits/scan)	7.3
MODIS-N TIR Data (megabits/scan)	3.2
MODIS-N Daytime Data (megabits/scan)	10.5
MODIS-T Daytime Data (megabits/scan)	12.6
MODIS-N # Scans/Orbit	5000
MODIS-T # Scans/Orbit	579
MODIS-N Daytime Data Rate (mbps)	8.9
MODIS-N Nighttime Data Rate (mbps)	2.7
MODIS-T Daytime Data Rate (mbps)	2.7
MODIS-N Orbital Ave Data Rate (mbps)	5.8
MODIS-T Orbital Ave Data Rate (mbps)	1.2
MODIS-N Daily Data Volume (gigabytes)	62.6
MODIS-T Daily Data Volume (gigabytes)	13.1
Total Daily Data Volume (gigabytes)	75.8
MODIS-N Volume (gigabytes) Level-1A	65.8
MODIS-T Volume (gigabytes) Level-1A	13.8
Total Daily Volume (gigabytes) @1A	79.6
MODIS-N Volume (gigabytes) Level-1B	113.6
MODIS-T Volume (gigabytes) Level-1B	23.1
Total Daily Volume (gigabytes) @1B	136.7
Total Daily Volume (gigabytes) @1A&1B	216.3

TABLE 5

MODIS Long-Term Archive Storage Requirements
(Gigabytes Per Day)

DATA PRODUCT	PRODUCT LEVEL					
	1A	1B	2/T	2/N	3	TOTAL
Navigation		18.7				18.7
Calibration		6.8				6.8
Spacecraft Ancillary	4.3					4.3
At-Satellite Radiances	75.3	104.7				180.0
Water-Leaving Radiances			10.1	4.0	6.6	20.7
Single Scattering Aerosol Radiances			8.2	2.6		10.9
Angstrom Exponents			0.3	0.4		0.8
Chlorophyll-A Concentrations (Case 1)			0.3	0.4	0.2	1.0
Chlorophyll-A Concentrations (Case 2)			0.0	0.0	0.2	0.3
Chlorophyll-A Fluorescence			0.3	0.4	0.2	1.0
CZCS Pigment Concentrations			0.3	0.4	0.2	1.0
Sea-Surface Temperature				1.1	0.2	1.3
Sea-Ice Cover				0.1	0.1	0.2
Attenuation at 490 nm			0.3	0.4	0.2	1.0
Detached Coccolith Concentration			0.1	0.1	0.2	0.5
Phycoerythrin Concentrations			0.3		0.2	0.5
Dissolved Organic Matter			0.3	0.4	0.2	1.0
Suspended Solids			0.3	0.4	0.2	1.0
Glint Field			0.3	0.4	0.2	1.0
IPAR			0.1	0.1	0.1	0.2
Ocean Cal Data Sets						0
Primary Production (Oceans)			0.3	0.4	0.2	1.0
Land-Leaving Radiances			2.7	14.4	2.8	20.0
Topographically Corrected Radiance			2.7	14.4	2.8	20.0
Vegetation Index				3.8	1.7	5.5
Polarized Vegetation Index				3.8	1.7	5.5
Land Surface Temperature				0.5	0.2	0.6
Thermal Anomalies				0.5		0.5
Evapotranspiration					0.1	0.1
Primary Production (Land)					0.1	0.1
Snow Cover				0.2	0.0	0.2
Spacial Heterogeneity (not sized here)						0
Land Cover Type					0.0	0.0
Bidirectional Reflectance, BRDF					0.0	0.0
Cloud Mask			0.3	1.6		1.9
Cloud Fraction					0.0	0.0
Cloud Effective Emissivity				0.1	0.0	0.1
Cloud-Top Temperature and Pressure				0.3	0.0	0.3
Cloud Optical Thickness (0.66 fm)				0.1	0.0	0.1
Cloud Particle Effective Radius				0.1	0.0	0.1
Cloud Particle Thermodynamic Phase				0.0	0.0	0.0
Aerosol Optical Depth (0.41 to 2.13fm)					0.0	0.0
Aerosol Size Distribution					0.0	0.0
Aerosol Mass Loading					0.0	0.0
Atmospheric Stability				0.1	0.0	0.1
Total Precipitable Water				1.7	0.0	1.7
Total Ozone				0.1	0.0	0.1
Browse		6.5	1.4	2.7	1.0	5.0
Metadata (Not sized here)						0
Ocean Discipline Subtotal (L-2/3)			21.8	12.1	9.5	43.4
Land Discipline Subtotal (L-2/3)			5.4	37.6	9.5	52.5
Atmosphere Discipline Subtotal (L-2/3)			0.3	4.0	0.0	4.3
Total	79.6	136.7	28.9	56.4	20.0	315.0

TABLE 6

Estimated MODIS Data Processing Requirements
(Lines of Code)

PROCESSING LEVEL	LAUNCH LOC	GROWTH LOC
Level-1A	25,000	25,000
Level-1B	25,000	30,000
Calibration/Monitor	72,000	144,000
Level-2 Ocean	12,000	24,000
Level-2 Land	40,000	80,000
Level-2 Atmosphere	20,000	40,000
Level-2 Shell	30,000	30,000
Level-2 Utility	40,000	80,000
Level-2 IDS Products	36,000	72,000
Level-3	30,000	60,000
Near-Real-Time	17,800	81,500
Subtotal	347,800	666,500
Supporting Software (validation)		552,000
Total		1,218,500

TABLE 7

MODIS Data Distribution
(Gigabytes Per Day)

FROM	TO	DATA DESCRIPTION	DATA VOLUME
CDOS	GSFC	All Level-0 Products	76
GSFC	MODIS Investigators	10% of Level-1A Products 50% of Level-1B Products 100% of Level-2 Products 100% of Level-3 Products	182
GSFC	Other Investigators	5% of Level-1B Products 10% of Level-2 Products 10% of Level-3 Products	17
GSFC	EDC	Level-1B for Land Products	41
GSFC	Langley Research Center	100% of Level-1B Products	137
GSFC	National Snow and Ice Data Center (NSIDC)	Level-1B for Snow and Ice Products	4

N92
12428
UNCLAS

S3-82
N92-12428
p.12

SAR DATA COMPRESSION: APPLICATION, REQUIREMENTS AND DESIGNS

J. C. Curlander and C. Y. Chang
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA

Abstract. The feasibility of reducing data volume and data rate is evaluated for the Earth Observing System (EOS) Synthetic Aperture Radar (SAR). All elements of data stream from the sensor downlink data stream to electronic delivery of browse data products are explored. This paper analyzes the factors influencing design of a data compression system including the signal data characteristics, the image quality requirements and the throughput requirements. The conclusion is that little or no reduction can be achieved in the raw signal data using traditional data compression techniques (e.g., vector quantization, adaptive discrete cosine transform) due to the induced phase errors in the output image. However, after image formation a number of techniques are effective for data compression.

1. Introduction

The Earth Observing System (EOS) is a joint program involving the National Aeronautics and Space Administration (NASA), the European Space Agency (ESA) and the National Space Development Agency (NASDA) [1]. Its prime objective is to provide long term monitoring of the earth as a system and quantitatively analyze the factors affecting global change. Four platforms (EOS-A, EOS-B, POEM of ESA and the NASDA platform) will be deployed, each carrying ten to twenty instruments selected to optimize the synergism resulting from simultaneous observations. Each platform is designed for a five year life cycle and will be followed by two identical platforms for a total fifteen year observation period.

In addition to the L and C band synthetic aperture radars (SARs) to be flown on the NASDA and ESA platforms respectively, a NASA sponsored SAR planned for a 1999 launch will be flown on a dedicated (Delta launched) spacecraft due to its unique characteristics [1]-[2]. The EOS SAR will operate at three frequency bands and four polarization channels similar to the SIR-C/X-SAR mission [3]. Table 1 shows the orbit and sensor characteristics of EOS SAR. The EOS SAR data will be acquired using a variety of swath and resolution modes for both strip and scanning data acquisition as shown in Table 2. The planned scenario is for the EOS SAR to collect data at an average data rate of 15 Mbps (with a peak data rate of 180 Mbps). The processor is required to operate at a throughput rate equal to the average data acquisition rate (with 50% margin) to generate the data products for delivery to the end users. Table 3 defines the various types of SAR data products. Because of the huge volume of signal data collected by the radar as well as the image data generated by the processor, efficient coding of these data would significantly decrease both the transmission and archive costs.

In this paper, we present study results on data compression for the EOS SAR applications. Section 2 discusses the SAR data characteristics with the communication system characteristics and constraints discussed in Section 3. Section 4 summarizes the performance of the evaluated data compression algorithms. Potential scientific applications and constraints of these techniques are presented in Section 5.

2. SAR Sensor and Data Characteristics

For any given sensor, the data characteristics establish the basis for the design of the data compression algorithm. The key parameters include the entropy, the rate distortion function and the stationarity properties of the data set. The entropy of the data determines the maximum compression ratio that can be achieved using a lossless data compression algorithm. Similarly, the rate distortion function, for a

given performance distortion criterion, determines the maximum compression ratio that can be achieved using a lossy data compression algorithm. Non-stationarity of the data statistics in the spatial and temporal domains imposes the requirement of adaptivity on the data compression algorithm.

For SAR signal data, the entropy is normally greater than seven bits per data sample for eight bit quantization based on a Gaussian distribution model. Previous studies have shown that a compression ratio of 3:1 (6:1) can be achieved at 12 dB (9 dB) signal-to-distortion noise ratio [4]. The degradation in image quality from this type of compression is quite severe due to distortion of the phase information required to form the image products. Compression at this stage would preclude all but the most qualitative science applications. The SAR signal data is processed into imagery using a two-dimensional matched filtering operation [5]. For a magnitude detected byte image product, the data is Rayleigh distributed with an entropy of approximately six to seven bits. Since the power of the return SAR echo is modulated by the two-way antenna pattern, the slant range attenuation and the varying resolution cell in the cross-track direction, the SAR data exhibits a wide dynamic range. Additionally, the target backscatter coefficient varies in both along-track and cross-track directions such that the stationarity is generally not valid for target areas greater than 10 Km^2.

The parameters used to characterize the SAR image quality include the resolution, sidelobe ratios and cross-channel relative phase error of the point target response functions as well as the image radiometric and geometric fidelity. A performance evaluation of the data compression algorithm should focus not only on the signal to distortion noise ratio but also on the resultant effects on these image quality parameters. Obviously, the effects of data compression on the inversion algorithms used for scientific analysis of the image products is the deciding factor as to the effectiveness of the compression operation. However, since these criteria are highly application dependent, we will only apply distortion measures to the intermediate data to which the data compression is applied.

3. Communication System Characteristics and Constraints

Figure 1 presents a functional block diagram of a digital communication system with source encoder (or data compressor), channel encoder (or error correction coder), modulator, demodulator, channel decoder and source decoder. In contrast to the source coding which is applied to remove redundancy from the source data, the channel coding is employed to improve the reliability of data transmission by inserting redundant data. In a conventional communication system, these components are designed and implemented independently. An efficient communication system design should consider the net compression ratio of the source data rate to the data rate transmitted through the communication channel since the channel effects can become significant for some data compression schemes. These schemes make the data more susceptible to bit errors and may not effectively provide any compression due to the overhead incurred by the required channel coding. From the end-to-end communication system point of view, the requirement should be set to maximize the number of bits per source data sample per unit bandwidth used in the analog communication channel.

There are three major segments in the communication system for the EOS SAR. The first one is from the platform via the TDRSS to the TDRSS ground receiving station at White Sands. The second one is from the White Sands ground receiving station to the designated data processing center(s). The third one is from the data processing center(s) to the end users, which is via the NASA science data network typically at a lower data rate (9600 bits per second) than the downlink.

For the data link from the platform via the TDRSS to the ground receiving station, there are two grades of services available: Grade II and Grade III services [6]. The Grade III service achieves a bit error rate of 10-5 for a 4.5 dB signal-to-noise ratio by employing a constraint length 7, rate 1/2 convolutional code modulated using QPSK. To achieve the required bit error rate, a channel coding has been employed that doubles the effective science data rate. Furthermore, the bit errors uncorrected by the convolutional code, will result in burst errors. In the Grade II service, the (255, 223) Reed-Solomon code is employed

as the outer code to correct these burst errors which improves the bit error rate to 10-8 (at the same signal-to-noise ratio) with an increase in the data rate of 14%.

For the EOS SAR, the requirement is for a bit error rate of 10-5 for the SAR signal data and 10-8 for the relatively low data volume auxiliary data. Given the channel link SNR = 4.5 dB, there may well be more efficient channel coding schemes than currently offered for downlink of the SAR data stream. For example, a high rate convolutional code combined with a multi-level, phase shift keying would be a good area of research to determine if the required link capacity could be reduced without data compression [7].

4. Data Compression Algorithms

In general, there are two classes of data compression algorithms [8]-[10]. One is the lossless coding algorithms used for applications that require exact reconstruction of the original data set. The other is the lossy coding algorithms used for applications where some level of compression noise is acceptable. It is worth noting that under special conditions some algorithms which are normally categorized as lossy may become lossless. In the selection of data compression algorithm, four factors need to be considered. They are the compression ratios, the compute facility available at both the transmitting and receiving stations, the reconstructed image quality and its sensitivity to bit errors. A final determination of the optimal algorithm will depend on the specific application requirements.

4.1 Lossless Coding Algorithms

The generally used lossless coding algorithms include Huffman coding and universal noiseless coding [8], [11]. The Huffman coding algorithm requires the knowledge of the probability distribution while the universal noiseless coding algorithm only requires the probability ordering of the source data. The probability ordering characteristics can be obtained by preprocessing the data samples. For SAR data, since the entropy is high (approximately 6 to 7 bits per sample for 8 bit quantization), the maximum compression ratio is limited to < 1.3. Given the addition of channel coding required to protect this compressed data from bit errors, the effective reduction using lossless coding does not justify the cost and complexity of the implementation.

4.2 Lossy Coding Algorithms

The lossy coding algorithms can be categorized into predictive coding, transform coding, vector quantizer, and a variety of ad hoc techniques [8]-[15].

The predictive coding is a relatively simple coding algorithm that results in a small compression ratio with reasonably good image quality [12]. Its major limitation is that it cannot compress the data below one bit per pixel. For most SAR applications, the quality of a reconstructed image using one bit per sample is unacceptable. To accommodate the non-stationarity property, the input data must be buffered to update the prediction coefficients on a frame by frame basis. Note that the predictive coding algorithm becomes lossless if the dynamic range of the prediction errors is retained, in which case the compression ratio is determined by the entropy of the prediction errors.

The adaptive transform coding is an algorithm capable of compressing the image data to any user specified compression ratio given that the associated image quality degradation is tolerable. Its major limitation is that it is computationally intensive and requires large buffers for both encoding and decoding. For most SAR applications, it generally yields an image quality better than other lossy coding algorithms. To accommodate the non-stationarity property, the class map which characterizes the block adaptivity must be updated every image frame. Figure 2 shows a Seasat Los Angeles image compressed by the adaptive discrete cosine transform coding algorithm with a 100:1 compression ratio.

The vector quantizer (VQ) is capable of producing good reconstructed image quality at high compression ratios. As compared to the adaptive transform coding algorithm, the primary advantage of the VQ algorithm is its simple decode procedure. The major drawback of the VQ is the complexity involved in the codebook training and data encoding. To reduce the encoding complexity, tree-searched schemes are employed such that the complexity only grows linearly rather than exponentially as the codebook size is increased. For SAR, the codebook must be updated every image frame or adaptive to the local data statistics using automatic gain control. Figure 3 shows a Seasat Beaufort Sea image compressed by a two-level tree-searched vector quantizer with a 16:1 compression ratio.

5. Potential EOS SAR Applications for Data Compression

There are a number of data system elements where the EOS SAR may utilize data compression. They include the downlink data stream, the primary data archive, and the image browse system.

5.1 Downlink of Data Stream

Spatial compression of SAR signal data is generally not feasible due to the phase fidelity required for the image formation matched filtering process. Implementation of a sophisticated, on-board data compressor which must include the SAR signal processor is a costly option that is not well accepted by the science community. There are two alternative techniques to achieve reduction in the downlink data rate. One approach is to reduce the overhead incurred by the channel coding scheme. This may be achieved by employing the high rate convolutional code combined with a multi-level, phase modulation scheme without the Reed-Solomon code as the outer code. The other approach is to employ a simple, adaptive data compression scheme, such as block floating point quantizer (BFPQ) which uses a fixed number of bits to quantize the data relative to a reference scale that is represented by additional data to characterize the global variation of data statistics. The latter approach has been successfully employed by the Magellan SAR system and will be used by SIR-C and EOS SAR.

For quick-look applications, a relatively simple on-board processor followed by a data compressor could be employed to fit the data within a low rate broadcast link (< 1 Mbps). For this quick-look application, a tree-searched vector quantizer is considered as a good candidate because it requires only a small workstation at the receiving stations for reconstruction of the compressed image data. Furthermore, its encoder can be implemented using relatively low cost, space qualified VLSI chips [16].

5.2 Primary Data Archive

The data set stored in the primary archive will be used by the end users for quantitative analysis which requires no loss in data information. Because of the speckle inherent in the SAR image data, only small compression ratio can be realized by lossless compressor. Using the basis that the data compression technique is only considered feasible if its implementation cost is lower than the savings from the archive storage capacity, a combination of predictive coding and universal noiseless coding appears to be a good candidate. The source data will first pass through a linear predictor. The prediction errors, which normally assume a smaller dynamic range than the source data samples and also exhibit the probability ordering characteristics, are then passed to the universal noiseless coder for removal of redundancy in the data. The implementation cost for the coding will be small since the technology for a custom hardware board is well proven [11] and little buffering capability is required.

5.3 Browse Data Products

The image browse system is designed for end users to quickly examine the image products that are routinely generated by the processor prior to delivery of high precision data products. The image data will be electronically transferred via a low data rate network, such as the NASA space physics analysis network (SPAN), to users with limited compute facilities available for reconstruction of compressed

image data. Since there is more compute power available in the primary data processing facilities, the encoding complexity is a less critical issue than the decoding. For browse applications, image quality and transfer time corresponding to compression ratio between 10:1 and 20:1 are adequate for quick-look analysis. The tree-searched vector quantizer meets all the above requirements

6. Summary

This paper summarizes a variety of factors influencing the feasibility of using data compression for the EOS SAR. In consideration of an EOS SAR data compression system, several factors have been evaluated: the data characteristics, the various system elements and the cost trade-off issue. Not discussed here but of key importance is the fact that the performance evaluation of any data compression algorithm must consider the induced distortion noise from the compression operation as well as the effects on the scientific inversion algorithms. The net compression ratio of the end-to-end communication system was considered with the conclusion that for an efficient communication system design, source coding, channel coding and modulation should be integrated into a single system. The compute facility available on both the transmitting and receiving stations is also a significant factor for algorithm selection. Assuming the image quality is acceptable, the net cost impact (i.e., cost savings from reduced channel link capacity and archive storage capacity minus implementation cost) is the final determining factor that will establish the feasibility of employing data compression for the EOS SAR system. This may be significant for the SAR due to the large volume of data and high data rates involved.

References

[1] *Earth Observing System: Reference Handbook*, NASA Goddard Space Flight Center, 1991.

[2] *Earth Observing System: Synthetic Aperture Radar Instrument Panel Report*, Vol. IIf, NASA, 1988.

[3] R. Jordan, B. Huneycutt and M. Werner, "SIR-C/X-SAR Synthetic Aperture Radar Systems," *Proceedings of the IEEE* (Special Issue on Spaceborne Radars for Planetary Observations), 1991.

[4] Read, C. J., Arnold, D. V., and Chabrias, D. M., "Synthetic Aperture Radar Image Formation from Compressed Data Using a New Computation Technique," *IEEE AES Magazine*, October 1988.

[5] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, John Wiley and Sons, 1991.

[6] *Tracking and Data Relay Satellite System (TDRSS) Users Guide*, Rev. 4, 1980, GSFC.

[7] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. on Information Theory*, Vol IT-28, No. 1, pp. 55-67, Jan 1982.

[8] Lynch, T., *Data Compression Techniques and Applications*, Van Nostrand, NY 1985.

[9] Jain, A., "Image Data Compression: A Review," *Proceedings of the IEEE*, Vol. 69, No. 3, pp. 349-389, March 1981.

[10] Chang, C. Y., Kwok, R., and Curlander, J., "Spatial Compression of Seasat SAR Imagery," *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 26, No. 5, pp. 673-685, September 1988.

[11] Rice, R., and Lee, J., "Real-time Data Compressor for EOS Class Mission," *Eascon '87*, October 1987.

[12] A. Habibi, "Comparison of the nth-order DPCM Encoder with Linear Transformation and Block Quantization Techniques, " *IEEE Trans. on Comm. Tech.*, Vol. COM-19, pp. 948-956, Dec 1971.

[13] P. A. Wintz, "Transform Picture Coding," *Proceedings of the IEEE*, Vol. 60, No. 7, pp. 809-820, July 1972.

[14] W. H. Chen and H. Smith, "Adaptive Coding of Monochrome and Color Images," *IEEE Trans. on Comm.*, Vol. COM-25, No. 11, pp. 1285-1292, Nov. 1977.

[15] Y. Linde, A. Buzo and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. on Comm.*, Vol. COM-28, No. 1, pp. 84-95, Jan 1980.

[16] W. C. Fang, et al, "A VLSI Design of Systolic Tree-Searched Vector Quantizer for EOS On-Board SAR Processor, " *Proceedings of GOMAC*, 1989.

Altitude	620 Km
Orbit	97.9° INCLINATION
Platform Power	5.8 KW PEAK POWER
Frequencies	L, C, X
Antenna Width	1.87 m (L), 0.48 m (C), 0.26 m (X)
Antenna Length	10.9 m
Polarization	HH, HV, VH, VV (L & C), HH OR VV (X)
Incidence Angle	15°-43.5° (SINGLE, DUAL-POL), 15°-32° (QUAD-POL)
Prf	1375 Hz - 2100 Hz
Chirp Bandwidth	20, 15, 10, 5, 1 MHz
Pulse Width	25, 34, 41, 50 usec
Data Rate	180 Mbps PEAK, 15 Mbps AVERAGE
Bits Per Sample	8, BFPQ
Peak Radiation Power	3.6 KW (L), 2.8 KW (C)

Table 1: EOS SAR orbit and radar characteristics.

Operation Mode	Resolution	Swath	Frequency	Polarization
Local, High Resolution Mode	20 - 30 m	30 - 45 Km or 80 - 90 Km	L & C or L & C & X	Quad-Pol or Single-Pol
Regional, Medium Resolution Mode (3 Scans)	80 - 110 m	150 - 240 Km	L & C or C & X	Single-Pol
Global, Low Resolution Mode (6 Scans)	250 m	~360 Km	L & C	Single-Pol

Table 2: EOS SAR operation modes.

Level 0:	Reconstructed unprocessed instrument data at full resolution.
Level 1A:	Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (i.e., platform ephemeris) computed and appended but not applied to the level 0 data.
Level 1B:	Level 1A data that has been processed to sensor units (i.e., radar backscatter cross-section). Standard SAR product.
Level 2:	Derived geophysical parameters (e.g., ocean wave height, soil moisture, ice concentration) mapped on some uniform time/space grid with processing parameters appended.
Level 3:	Geophysical data mapped on uniform space-time grid scales, usually with some completeness and consistency properties (e.g., missing points interpolated, complete regions mosaicked together from multiple orbits)
Level 4:	Model output or results from analysis of low-level data (i.e., geophysical data not measured by the instruments but derived from instrument measurements).

Table 3: SAR data product level definitions.

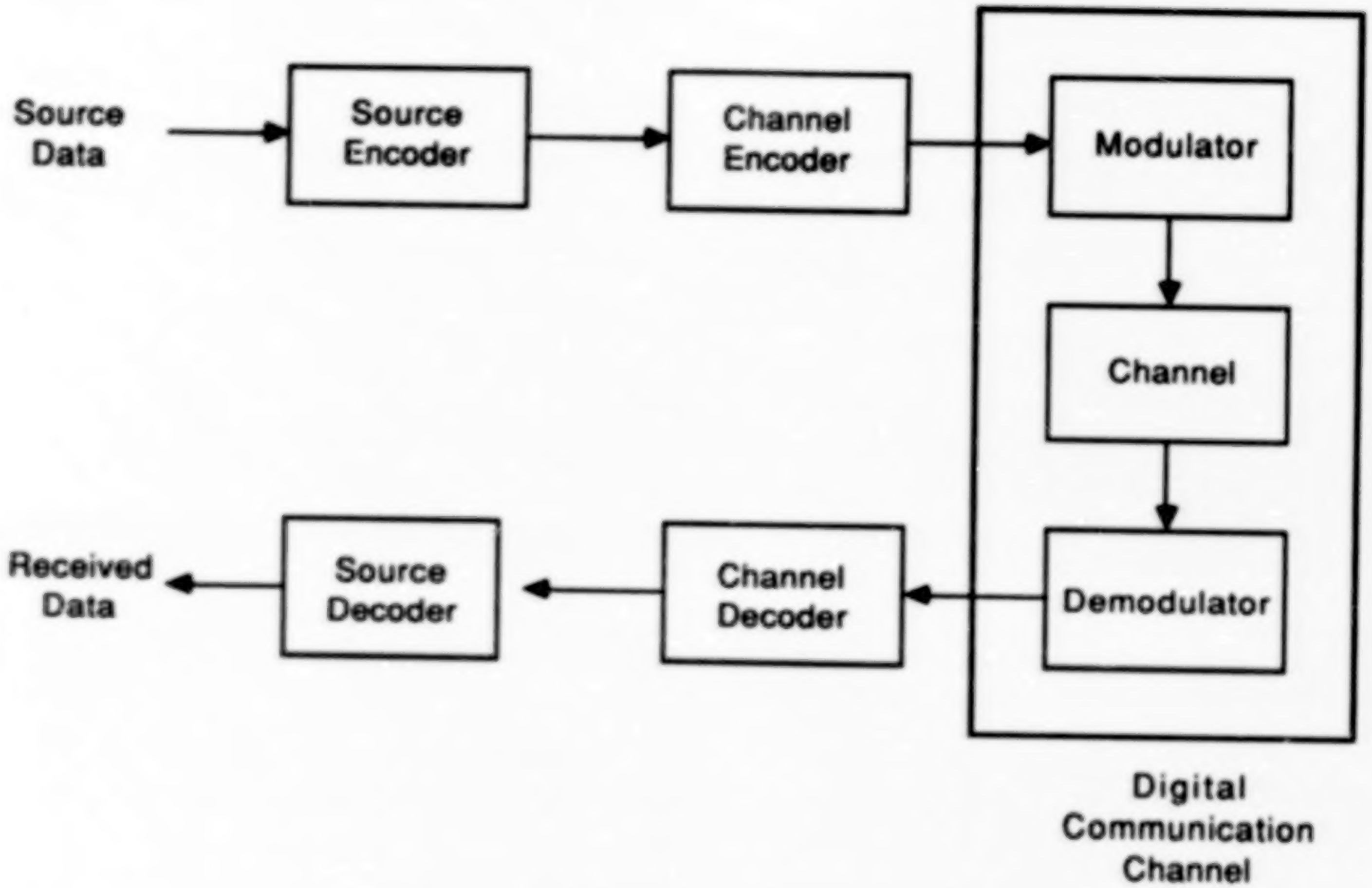


Figure 1: End-to-end communication system.

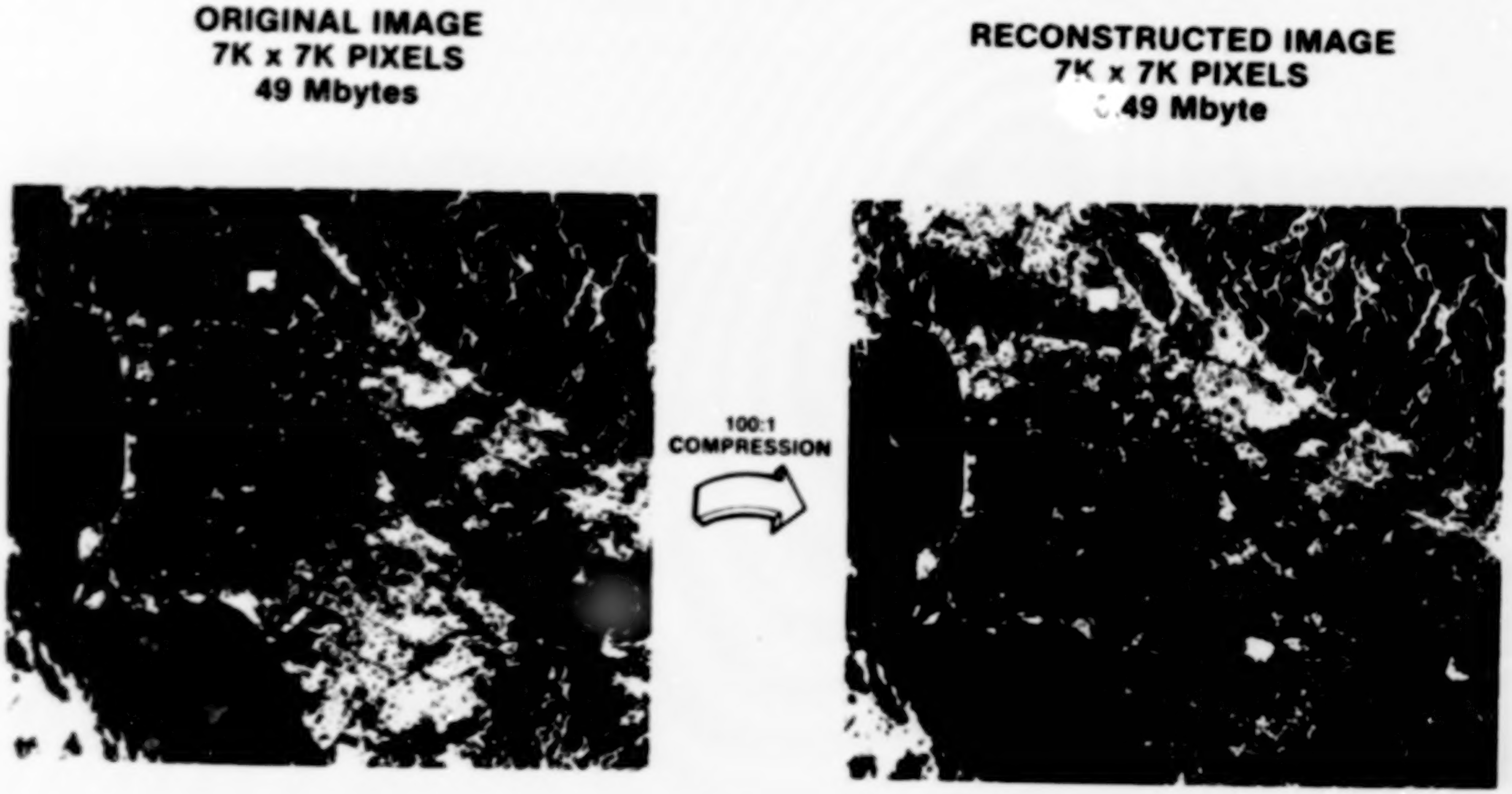


Figure 2: Compression of SAR imagery using adaptive discrete cosine transform algorithm

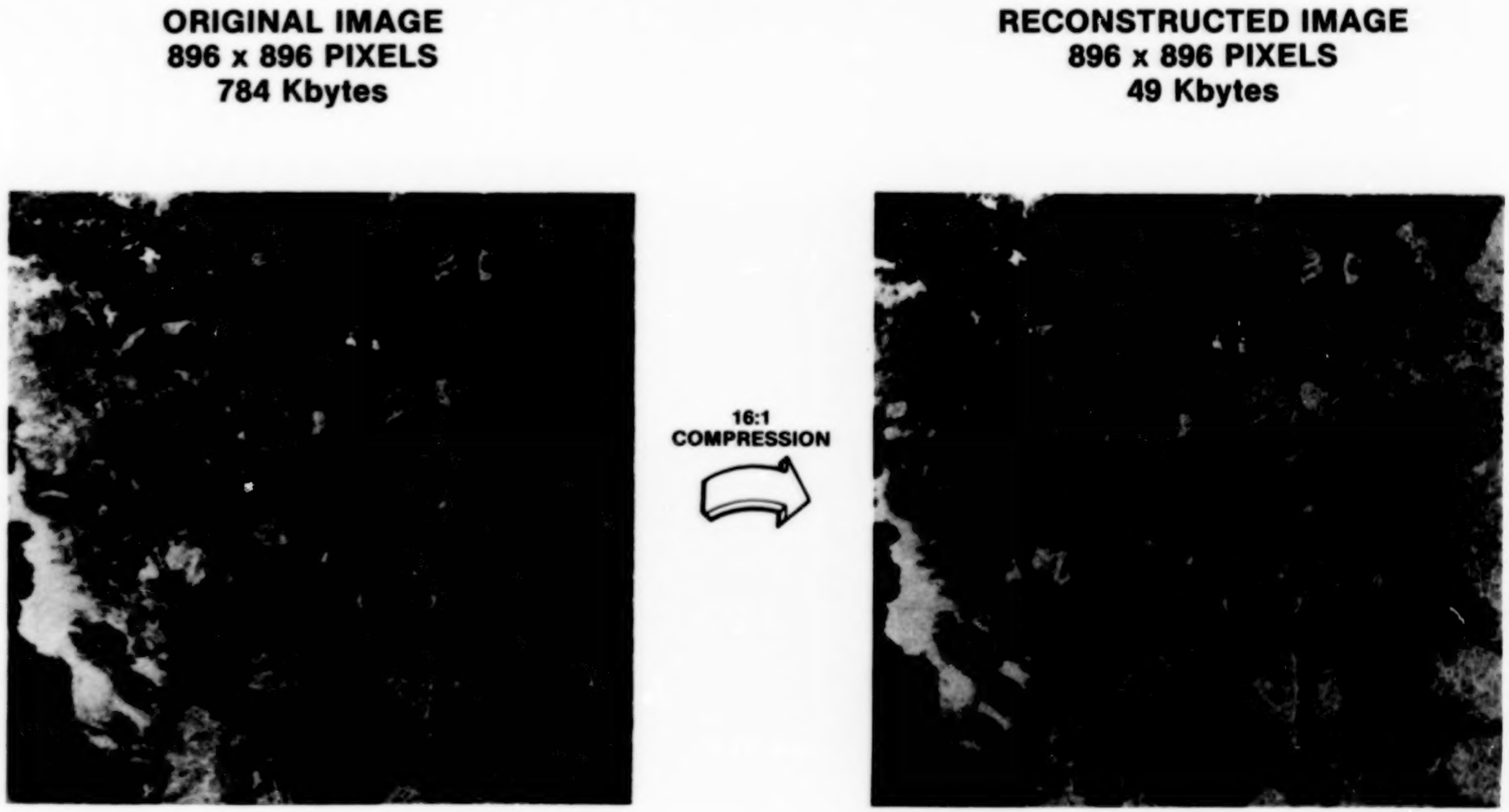


Figure 3: Compression of SAR imagery using two-level tree-searched vector quantizer

N92
12429
UNCLAS

S4-82
N92-12429
37866
p. 14

SCIENTIFIC REQUIREMENTS FOR SPACE SCIENCE DATA SYSTEMS

Raymond J. Walker
Institute of Geophysics and Planetary Physics
University of California
Los Angeles, CA 90025

Abstract. In the 1990's space plasma physics studies will increasingly involve correlative analysis of observations from multiple instruments and multiple spacecraft. The solar terrestrial physics missions in the 1990's will be designed around simultaneous observations from spacecraft monitoring the solar wind, the polar magnetosphere and the near and distant magnetotail. Within these regions clusters of spacecraft flying in formation will provide observations of gradients in the plasma and field parameters. Planetary plasma studies will increasingly involve comparative magnetospheric studies. No single laboratory will have the expertise to process and analyze all of the different types of data so the data repositories will be distributed. Catalog and browse systems will be required to help select events for study. Data compression techniques may be useful in designing the data bases used for selecting events for study. Data compression on board the spacecraft will be necessary since instrument data rates will be much larger than available telemetry rates. However, considerable care will be necessary to avoid losing valuable data when applying data compression algorithms.

1. Introduction

Space physics is a wide ranging discipline. It includes solar physics, heliospheric physics (the solar wind and interplanetary magnetic field), the physics of the magnetosphere, the physics of the ionosphere and the interaction between the plasmas in these regions. In addition space physicists are interested in that part of planetary science having to do with the interaction between the solar wind, planets, their moons, magnetospheres and ionospheres.

In this report we will discuss the requirements that studies of space plasmas place on the data systems. We will concentrate mainly on in situ data from spacecraft although many of the requirements are valid for ground based observations as well. The emphasis will be on studies that involve tensor time series data however many of the requirements are valid for remote sensing observations also. One of the main purposes of this volume is to acquaint computer professionals interested in data compression with the data problems encountered by scientists using space derived data. The approach in this paper will be to discuss the requirements on the entire data system from the perspective of a space scientist without trying to detail all of the areas where data compression could be useful. Hopefully this will start a dialog between the two communities which will help us define those areas where data compression techniques will be most applicable.

First we will consider a specific example of space physics research in the 1990's. The case we will examine is a study of the bow shock of Venus which was conducted by using observations from the Galileo spacecraft. We will examine the Galileo magnetometer observations and show how the results obtained in this study will lead to other studies which place requirements on the data system infrastructure. Next we will expand our view by considering the demands that the missions of the 1990's will place on the data systems. In particular we will consider the International Solar Terrestrial Physics Program. This international multispacecraft mission will be the prime project in solar terrestrial physics in the 1990's and will be the main driver for data activities in space plasma physics. Next we will examine the concepts currently being considered to solve some of the data problems in space plasma physics. We will do this by considering the distributed approach in space data management used by the Planetary Data System. Finally, we will briefly consider the applications where data compression has been used in space physics and will consider some of the concerns which arise in the science community whenever the use of data compression is suggested.

2. The Search for Intermediate Mode Shocks

2.1 What is an Intermediate Mode Shock?

Just as a hydrodynamic shock in a neutral gas converts a supersonic flow to a subsonic flow, a magnetohydrodynamic (MHD) shock in a plasma converts a flow which exceeds one of the phase velocities of the plasma to a velocity below it. In contrast to a neutral gas which has just one characteristic velocity, the sound speed, an MHD plasma has three speeds corresponding to three wave modes. They are the fast compressional mode, the slow compressional mode and the intermediate mode. The fast and slow mode waves are compressional (i.e. the magnetic field changes its magnitude as the wave propagates) while the intermediate wave is a shear wave in which the magnetic field changes direction but not magnitude. The changes in the parameters across a shock can be found by solving the Rankine-Hugoniot relations which express the conservation of mass, momentum and energy plus Maxwell's equations (Gauss' Law and Faraday's Law). These equations have six solutions (e.g. [1]) and it is useful to classify the shocks by the relationship between the flow velocities normal to the shock and the phase velocities of the MHD wave modes. Class 1 flows are faster than the fast velocity, class 2 flows are sub-fast speed but super intermediate speed, class 3 flows are sub-intermediate but super slow and class 4 flows are sub-slow speed. Thus the six types of shocks are (1,2) shocks in which the flow goes from super fast to sub-fast but super intermediate, (1,3) shocks which go from super fast to sub-intermediate but super slow, (1,4) shocks which go from super fast to sub-slow, (2,3) shocks which go from sub-fast but super intermediate to sub-intermediate but super slow, (2,4) shocks which go from sub-fast but super intermediate to sub-slow and (3,4) shocks which go from sub-intermediate but super slow to sub-slow.

It was long believed that only two of these solutions could exist in nature, the (1,2) shocks or fast shocks and the (3,4) shocks or slow shocks [2]. Both of these types of shocks have been observed in nature. The most famous example of a type (1,2) shock is the Earth's bow shock while slow shocks (3,4) are found in the Earth's magnetotail. Types (1,3), (1,4), (2,3) and (2,4) shocks are called intermediate shocks. Recently both theory and numerical simulation have suggested that these shocks too can exist [1,3].

Fast and slow mode shocks change the magnitude of the component of the magnetic field in the shock plane but do not change its sign. In an intermediate shock the component of the magnetic field along the shock surface must change sign across the shock [1]. There is only a small range of upstream flow conditions for which an intermediate shock can exist. For (1,3) or (1,4) shocks at $\beta < 1$ (β is the ratio of the plasma pressure to the magnetic pressure) , the upstream flow must have $1 < M_A < 2$ (the Alfven Mach number $M_A = v/c_A$ where the Alfven speed $c_A = B/(4\pi\rho)^{-1/2}$ with B the magnitude of the magnetic field and ρ the mass density). As β increases, the cutoff occurs for smaller M_A. The normal to the shock must be nearly along the magnetic field (such shocks are called parallel shocks). When the sound speed ($c_s = \gamma\, p/\rho$ where $\gamma = 5/3$ is the polytropic index and p is the pressure) is larger than c_A intermediate shocks of type (1,3) or (1,4) cannot exist but (2,3) and (2,4) shocks can. It is expected that shocks of types (1,3) and (1,4) might be attached to the fast mode bow shock while types (2,3) and (2,4) shocks will separate from it.

2.2 Galileo Observations

The Galileo spacecraft flew by Venus on February 10, 1990 as part of its voyage to Jupiter. The spacecraft approached Venus from the nightside on a trajectory which was nearly parallel to the expected position of the bow shock. Figure 1 shows the Galileo trajectory on the inbound leg near Venus. A model bow shock has been included. Since Venus has at most a very small intrinsic magnetic field the bow shock is very close to the surface of the planet near noon. The letters A-F indicate pairs of bow shock crossings. For these crossings on the flanks of the magnetosphere the magnetic field was nearly parallel to the expected shock normal. Thus this is a good region to look for intermediate shocks.

Figure 2 shows magnetic field observations from Kivelson et al., [4]. The three components of the field are plotted in Venus Sun Orbit (VSO) coordinates (x is toward the Sun, y is towards dusk and z is positive northward). The shocks can most easily be seen as sudden changes in the magnetic field magnitude in the bottom trace. The times between shock crossings are shaded. In this example we are mainly interested in the interval E between about 0334 UT and 0343 UT. This is shown in higher resolution in Figure 3. Here the traces in VSO coordinates are at the bottom of the figure as are simultaneous observations from one component on the Pioneer Venus Orbiter (PVO) spacecraft. The top panels show the Galileo magnetic field in shock normal coordinates with (I) along the direction of maximum variation and (K) along the shock normal direction while (J) completes the right hand system and lies in a plane perpendicular to the plane which contains the upstream and downstream vectors. The outbound shock crossing is at 03:43. Prior to that the field in the two transverse components rotates through nearly 180°. *Kivelson et al.*, [4] point out that this is consistent with either a fast (1,2) shock followed by a (2,3) intermediate shock or a (1,3) intermediate shock.

2.3 The Next Steps in the Study of Intermediate Shocks

The observations above are consistent with the 0343 UT event being an intermediate shock. However much more analysis will be required to establish that unambiguously. First we must establish that this is indeed a shock. Here observations from the plasma instrument and the plasma wave instrument on Galileo must be examined. The observations from the plasma instrument will help us determine if shock related heating has occurred. The plasma wave observations will help us determine if broad band radiation associated with a shock crossing is present. The addition of plasma data will give us the flow velocity, the density and the pressure and we will be able to calculate the critical parameters c_S, c_A and ß. With this we can determine whether or not these events are in the regime in which intermediate shocks can exist.

Even if all the evidence supports our suggestion that this is an intermediate mode shock we will still need to examine more data. We will need to investigate the other Galileo shocks looking for other examples of possible intermediate mode shocks and to try to determine empirically when intermediate mode shocks can occur. PVO also provides a potential source to be probed for evidence of intermediate shocks. The Earth's bow shock, too, is a possible source of data on intermediate shocks. The 9 years of data from the International Sun Earth Explorers (ISEE) spacecraft and data from IMP-8 should be examined. It is possible that the event identified above isn't an intermediate shock at all. For instance it could be a rotational discontinuity in the solar wind which reached the bow shock just as Galileo did. Examples with data from more than one spacecraft will be very valuable. With data from one spacecraft in the solar wind and one at the bow shock this possibility can be eliminated. In addition we can look for intermediate shocks propagating in the solar wind.

From a data system perspective, the most important lesson from this example is that modern space plasma physics requires data from a variety of instruments on a spacecraft and frequently from many spacecraft. Often that data must be from several instruments on several spacecraft simultaneously. Getting this data to the scientists in a timely manor is one of the major problems facing the designers of space science data systems. Indeed one of the major new missions in space physics, the International Solar Terrestrial Physics (ISTP) Program is based on this concept of using simultaneous observations from many instruments and many spacecraft. We will discuss it in the next section.

3. Multispacecraft Missions

The very nature of the magnetosphere requires that it be probed by multiple spacecraft simultaneously. The magnetosphere is vast and highly dynamic. Spacecraft observers are required to infer the dynamics of this system from time-series observations constrained to the spacecraft's trajectory. Without multiple point measurements they simply cannot tell what is happening in the rest of the system.

3.1 The International Solar Terrestrial Physics Program

A major question in magnetospheric physics is to understand the flow of energy and momentum through the solar wind, magnetosphere and ionosphere system. ISTP is a cooperative venture between NASA, the European Space Agency (ESA), and the Japanese Institute for Space and Astronautical Science (ISAS) to study this problem. In addition there are a number of associated missions from the Space Research Institute (IKI) of the USSR Academy of Sciences.

In ISTP, the Solar Heliosphere Observatory (SOHO) will remotely observe the Sun and make in situ observations of the composition of the solar wind from the L1 Lagrangian point. The Wind spacecraft will observe the solar wind and will provide the solar input to studies of the interaction of the solar wind with the magnetosphere. It, too, will be in a halo orbit at the L1 point. The Polar spacecraft will investigate the polar magnetosphere and remotely sense the auroral zone. The ESA Cluster mission will provide four spacecraft flying in a tetrahedral formation with identical instruments to measure gradients in the polar magnetosphere. The Japanese Geotail spacecraft will probe both the distant magnetotail out to $220R_E$ and the near Earth magnetotail. ISTP also will utilize observations from several associated missions. These include the Air Force/NASA CRRES satellite which monitors the inner magnetosphere out to about $6R_E$. Two Soviet missions may also contribute to ISTP. One of these Interbol will consist of two spacecraft each with a small subsatellite. One pair of spacecraft will be in polar orbit while the other pair will probe the tail out to about $35R_E$. Another planned Soviet mission is Regatta. Project Regatta comprises a system of four to five small space laboratories. The first of these is planned for the near earth tail with apogee at about 8 to $10R_E$. Later a polar Regatta spacecraft may join the ESA Cluster mission. It would orbit near the Cluster at about 10 times the tetrahedral spacing. Later in the decade additional Regatta spacecraft may join the ISTP group. Please see *Farquhar* [5] for more information on the ISTP spacecraft and their planned trajectories.

In addition to the spacecraft, the ISTP mission also will include coordinated ground observations from magnetometer chains and auroral radar. Finally ISTP will have a major program of theory and simulation investigations. Large scale models of the interaction between the solar wind, the magnetosphere and the ionosphere will be used to help organize these observations and the observations will help us test and refine the models.

3.2 Data System Requirements

Each of the ISTP spacecraft will have a complement of space plasma and fields instruments. The key element of ISTP is that much of this data will have to be analyzed together in a coordinated fashion. The major data system driver in space physics in general and solar terrestrial physics in particular will not be the volume of data but the number of sources of data. The instruments on these spacecraft are very sophisticated and require expert interaction to produce usable data. Thus the data system supporting the ISTP mission must be distributed. The data and the scientists processing it are closely linked. The ISTP scientists are planning to work together on studying in detail magnetospheric events. To accomplish this they will need some sort of browse system to help select events (they call this a the key parameter system). When ISTP is in full operation there may be several groups of scientists studying several events simultaneously. In addition to being able to use the browse systems to help select the events, they will also need to be able to locate the data required for detailed study and to access it.

4. Planetary Data in the 1990's

In the proceeding sections we have examined some of the demands that space physics research in the 1990's will place on data system activities both by considering a specific research example and by considering the problems of the major mission in the field. Now we would like to consider one further

example. In this section we will consider the data system requirements of that part of space physics concerned with the planets and how the NASA Planetary Data System is trying to address those needs.

When discussing planetary science it is important to remember that you can't study just one part of planetary science in isolation. The disciplines and sub-disciplines are linked by physical processes. For example if you want to determine whether Mars and Venus have electrically conducting cores and hence dynamos you will need to study the solar wind. Since both planets are at best weakly magnetized you need to first understand the effects of the solar wind in inducing a magnetosphere before you can determine the extent of any intrinsic magnetic field and learn about the precesses within the planet that create it.

Studies of the jovian magnetosphere require an understanding of the physics and chemistry of the surfaces, and atmospheres of the moons as well as plasma physics. For instance the Voyager observations in Jupiter's magnetosphere demonstrated that much of the plasma has its origin at the moon Io. We now believe that charged particles from the magnetosphere remove neutral particles from the surface and atmosphere of Io by a process called sputtering. (The neutrals originally came from ioian volcanoes.) These neutrals are ionized by electron impact ionization or charge exchange and form a plasma. This then is the plasma that interacts with Io and fills the magnetosphere.

Just as was the case in solar terrestrial physics, studies of the planets frequently require data from more than one instrument on a spacecraft and the data is frequently widely distributed at the laboratories where the scientific expertise in found. In addition in planetary science comparative studies involving observations from more than one planet are becoming increasingly important. In planetary science archival studies also are important. There will be no new in situ data from Uranus or Neptune for a very long time. The next Saturn data is over a decade away as is the next particles and fields data from Venus. Data from some new planetary missions is being archived immediately. For instance the Magellan mission has provided archival data to the scientific community from the beginning.

4.1 The Planetary Data System

The NASA Planetary Division has tried to address the data needs of the planetary science community by forming the Planetary Data System (PDS). PDS was founded on the principle that "the data repositories which work best are those in which data are managed by scientists who are actively engaged in research" [6]. PDS was charged to "provide the best planetary data to the most users forever!" [*McMahon, personal communication, 1991*].

Since planetary science is multi-disciplinary and since the data and the expertise are widely distributed, PDS is a distributed system. There are six science nodes, the Rings Node at Ames Research Center, the Imaging Node at the USGS in Flagstaff Arizona, the Small Bodies Node at the University of Maryland, the Geosciences Node at Washington University, the Atmospheres Node at the University of Colorado and the Plasma Interactions Node at UCLA. Since planetary science is too broad for any one institution to have all of the required expertise each of the Nodes has subnodes which provide expertise on a specific scientific instrument or data type. PDS is managed from a Central Node at JPL and they maintain a technology development and testing laboratory. Finally the Navigation and Ancillary Information Facility (NAIF) at JPL acts as a Node for spacecraft trajectory, attitude and pointing data. PDS is responsible for obtaining the data for archiving, making sure it is of high quality and assisting the scientific community with data problems. PDS deposits all of its data in the National Space Science Data Center (NSSDC) for permanent archiving.

Figure 4 shows the projected planetary data archives between now and 1997. By 1997 the PDS archives will total about 2500 GB. Throughout this decade it will grow at a rate of about 400--500 GB per year.

4.2 The Plasma Interactions Node

The Planetary Plasma Interactions Node (PPI) of PDS is responsible for planetary particles and fields data. It is responsible for data relating to plasma physics in planetary systems. This includes the interaction of the solar wind with planetary magnetospheres, ionospheres and surfaces. Also of interest are the interactions of magnetospheric plasmas with the satellites and rings within planetary magnetospheres. These interests overlap those of other PDS nodes and close working relationships are maintained with the Atmospheres Node, as well as the Small Bodies Node and the Rings Node. The PPI Node has subnodes at the University of Iowa, the Goddard Space Flight Center as well as a separate Inner Planets Subnode at UCLA.

The specific goals of the PPI Node include helping to assure that high quality and usable data are available to the scientific community, helping scientists to determine the availability of data, helping them select the data needed for a specific study, helping them access that data and helping them with the analysis of the data.

The PPI Node uses several approaches to assure that high quality and usable data are available to the community. Foremost among these approaches is the peer review. All data submitted to PDS is reviewed by a panel of scientists and technicians prior to its formal release to the scientific community. The data peer review is analogous to the review of papers for publication in a journal. Indeed the entire process of ingesting data into PDS is similar to that of submitting a paper to a journal. The peer review checks both the science data and the metadata describing the science data. The metadata are maintained in the PDS Catalog. It includes descriptions of the spacecraft, the instrument, the data processing and most importantly known sources of contamination. In addition the catalog contains information about the quality of the science data. When a scientist orders data from PPI, PDS or the NSSDC the data are documented with PDS Labels. These labels include information on the quality of the data. Finally to assure that the data are adequately preserved PDS pioneered the development of the concept of placing the data on CDROM.

To help scientists locate the data, PDS and PPI use the catalog system. The high level PDS catalog points to large collections of data while the detailed level catalog is essentially an inventory of all of the data holdings and helps scientists to locate subsets of the data.

The catalogs also help a user select data. The detailed level catalog provides information with a granularity of one hour. In addition the PPI Node has developed a system to browse the PPI data archive. The browse data consists of an averaged subset of the full resolution data. It is maintained on-line all of the time and can be displayed graphically. The software to access the browse data and display it is based on a client server architecture. The front-end of this system can be distributed to assure rapid access to the data. Figure 5 shows a typical graphics display from the browse system. The user can design the display interactively.

The PPI system is based on a file management system which uses a relational data base management system. Figure 6 shows the schema for this file management system. Most importantly the tables contain the information required to build the displays in the browse system (Group Table) and information on the status (Status Table) of the data (i.e. the path to the data and whether it is on-line or off-line etc.). With this information the PPI Node can help users access the data and order it.

The order data subsystem of the PPI Node uses the file management tables in Figure 6 to help a user place an order for data. It uses the file management tables to locate the data, fills the order if the data is already on-line or schedules moving the data on-line if it is not. If orders are relatively small they are filled directly by the PPI Node. Larger orders are routed to the NSSDC.

Finally PPI Node supports a number of data analysis packages. These include the Interactive Data Language (IDL) and the UCLA Data Flow System [7]. PPI will also provide users with access to both

theoretical models and simulations of planetary plasma processes. Most importantly PPI maintains a group of experts on various fields and particles data types who are available for consultation.

5. Data Compression and Space Physics

We have seen that in the 1990's space physics will increasingly involve correlative analysis of data from multiple instruments and multiple spacecraft. That data will be distributed because the people who know about the data are distributed. Finally there will be an increased use of both theoretical and empirical models to help us organize these observations and to help promote understanding.

How can data compression techniques help? This is the question that the computer professionals working in this field and space physicists will have to work together to answer. In this section we will discuss a few areas where data compression may be useful. The list in certainly not exclusive. We will also consider the problems involved with using data compression techniques.

It seems fairly clear that selecting the data for analysis will take on new importance in the 1990's. Before starting on a lengthy study scientists will want to assess whether the data needed are available. When selecting between two events for study they will be interested for instance in knowing for which event solar wind data are available, or whether auroral images are available. They will want to know where other spacecraft were located in the magnetosphere. Thus we believe that browse systems will take on increased importance. Being able to look at subsets of the data quickly will help in this selection process. Speed of access is very important for browse data. Researchers don't want to spend too much of their time in the selection process. Therefore the browse data should be on-line. This makes browse data a very good candidate for data compression. Since the user can always go back to the full resolution data when they conduct the detailed study, the browse data is also a likely candidate for lossy compression.

Some data compression is already being planned for instruments for future missions. The data rates of modern instruments have increased faster than the available telemetry. For some of the experiments the instrument data rate is as much as 20 to 40 times that which can be telemetered. Since the data rates of the instruments are closely coupled with the science, data compression is an attractive way to get the data back to Earth. Consider, for example, the magnetometer experiment on the ISTP Polar spacecraft. The minimum rate of data return is 10 vectors/s. Unfortunately this rate cannot be maintained by the allocated spacecraft telemetry. Here data compression by about a factor of four is required. A second differencing algorithm is being developed for use on the spin plane components. A second differencing algorithm will work on a spinning spacecraft like Polar since most of the signal is a sinusoid. Another limitation of the choice of the compression algorithm is that the on board processor must be able to carry out the compression in the time available with the available memory. Many powerful data compression algorithms have been rejected because they require more resources than are available on the spacecraft. So far the second differencing approach for the magnetometer is the only algorithm which will both provide the required compression and is fast enough to keep up with real time data.

The data compression being studied for Polar is lossless. This brings us to one of the major concerns which space physicists have when considering data compression algorithms. Instruments are designed to provide the data required to study a given phenomenon or set of phenomena. The instruments are carefully designed to provide the required measurements. Every bit is important for some potential study and scientists are reluctant to give up bits for data compression. Therefore lossy data compression is looked on with a great deal of suspicion. The computer professionals working on data compression techniques for space physics data will have to demonstrate that they aren't asking the scientists to give up science for compression.

Acknowledgments

The assistance of J. D. Means and D. Kloza in describing the efforts to apply data compression to data on board the ISTP spacecraft is gratefully acknowledged. The assistance of Todd King and Steven Joy with the PDS section of the paper is also gratefully acknowledged. This work was supported by the Jet Propulsion Laboratory under contracts 958694 (Galileo) and 959026 (PDS).

References

[1] Kennel, C. F., R. D. Blandford and P. Coppi, MHD intermediate shock discontinuities,. Part 1, Rankine-Hugoniot conditions, J. Plasma Physics, 42, 299, 1989.

[2] Kantrowitz, A. and H. E. Petschek, MHD characteristics and shock waves, in Plasma Physics in Theory and Application, W. B. Kunke, Editor, McGraw Hill Book Company, N. Y., 1966.

[3] Wu, C. C., On MHD intermediate shocks, Geophys. Res. Lett., 14,668, 1987.

[4] Kivelson, M. G., C. F. Kennel, R. L. McPherron, C. T. Russell, D. J. Southwood, R. J. Walker, C. M. Hammond, K. K. Khurana, R. J. Strangeway and P. J. Coleman, Magnetic field studies of the solar wind interaction with Venus from the Galileo flyby: First results, Science (submitted), UCLA Institute of Geophys. and Planet. Phys., Publication #3462, 1991.

[5] Farquhar, R. W., Handbook on trajectories, mission design, and operations for the spacecraft of the solar-terrestrial science project, The Inter-Agency Consultative Group for Space Science, November, 1990.

[6] Bernstein, R., S. S. Alexander, R. Arvidson, E. Olen, C. Goad, E. J. Groth, D. A. Landgrebe, R. Legeckis, R. L. McPherron, H. Tananbaum, D. L. Turcotte and T. H. Vondar Haar, Data management and computation. Volume 1: Issues and Recommendations, National Academy Press, 1982.

[7] King, T. A., and R. J. Walker, The UCLA Data Flow System, UCLA Institute of Geophys. and Planet. Phys., Publication #3522, 1991.

GALILEO TRAJECTORY IN VENUS-SUN-ORBIT COORDINATES

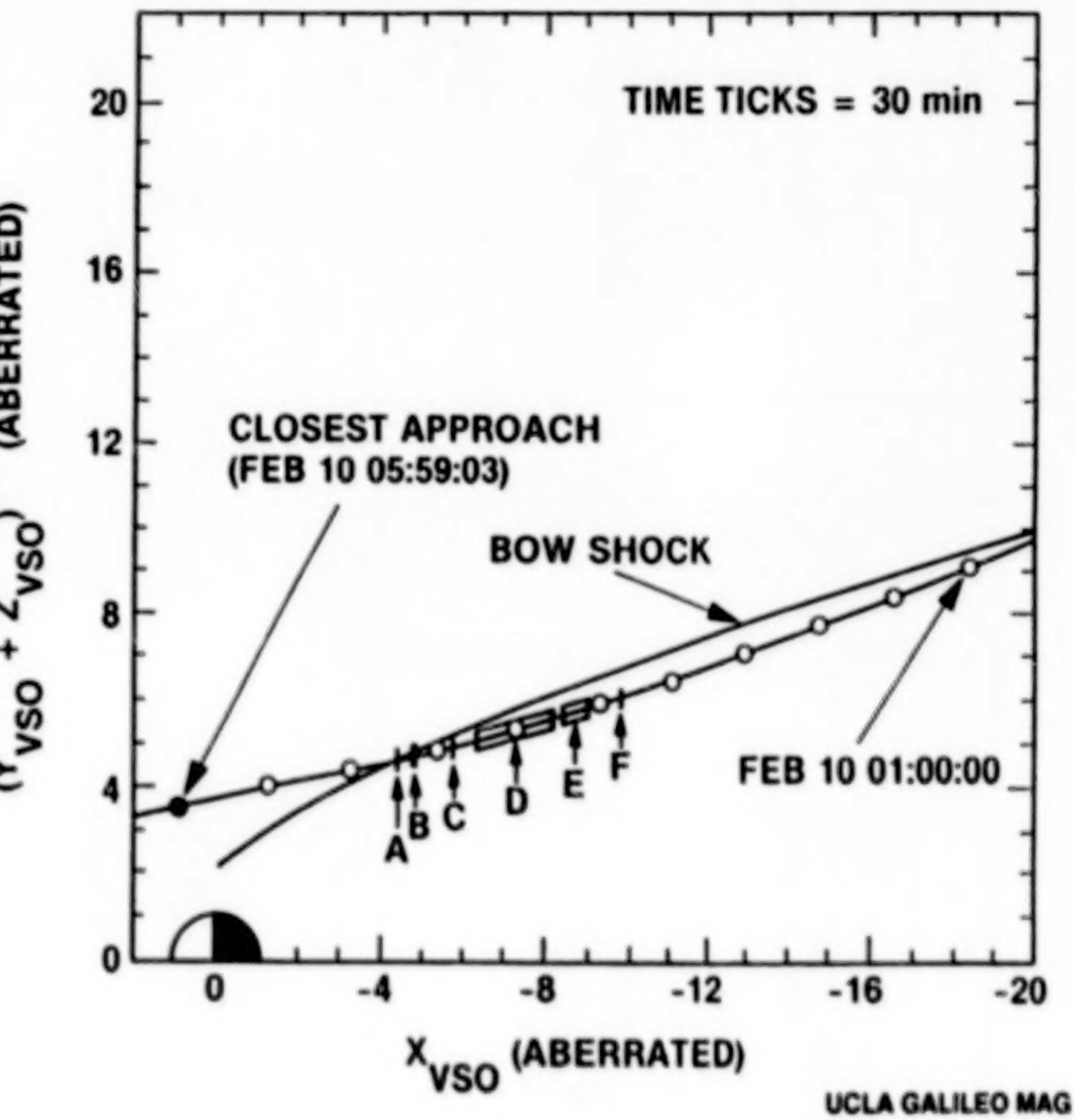


Figure 1. The Galileo trajectory near Venus in aberrated coordinates [4]. This view gives the trajectory in the plane of the spacecraft in terms of the distance along the solar wind aberrated planet-sun line and the perpendicular distance from that line. A model of the shock location is shown and the pairs of shock crossings (from upstream to downstream and then downstream to upstream) are labelled A-F.

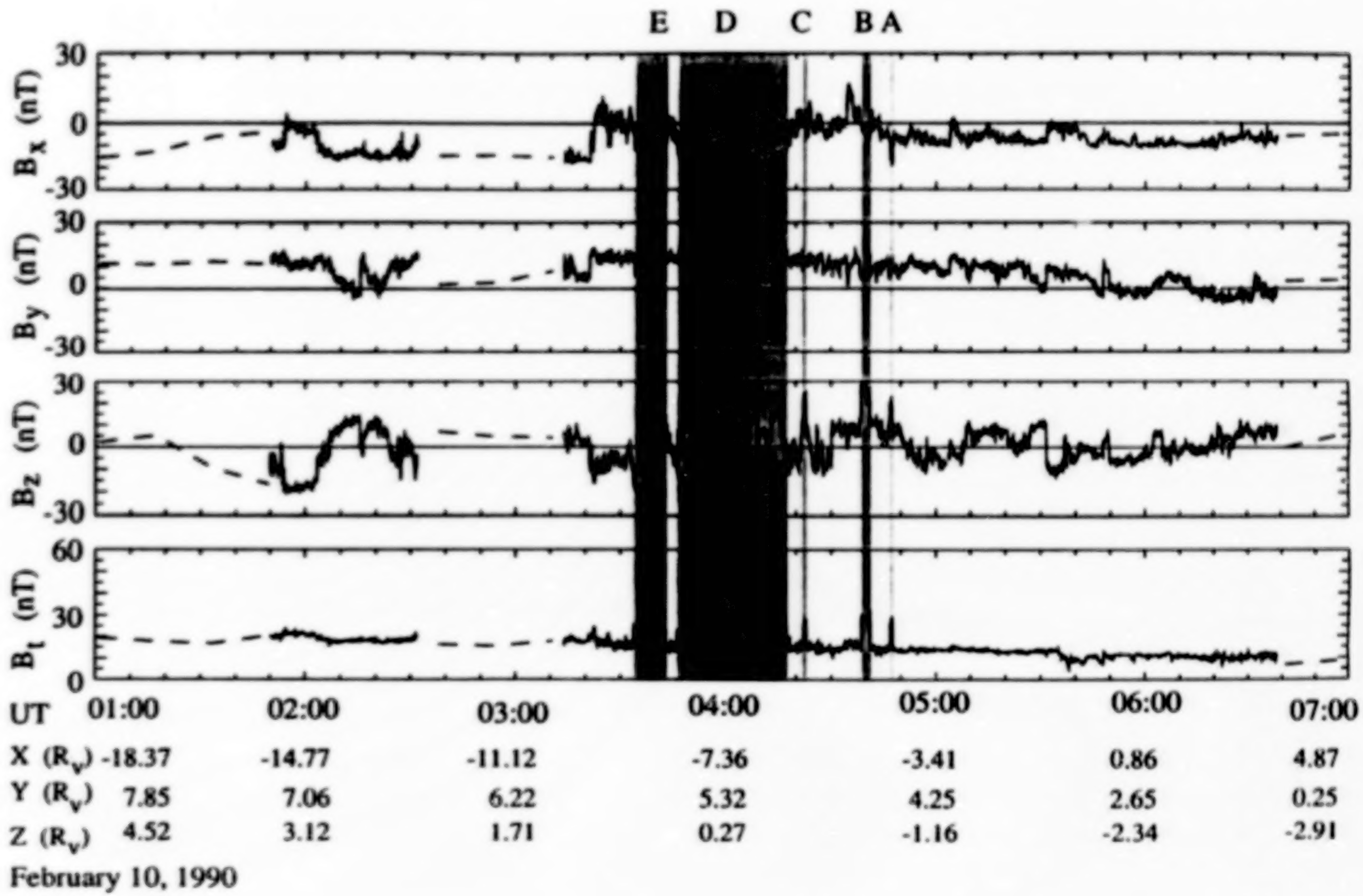


Figure 2. Magnetic field components and total field in VSO coordinates [4]. The shock crossing intervals in Figure 1 have been shaded. The gaps in the high time resolution data are filled in by using "optimal average" data taken on the spacecraft with 16 minute resolution (dashed lines).

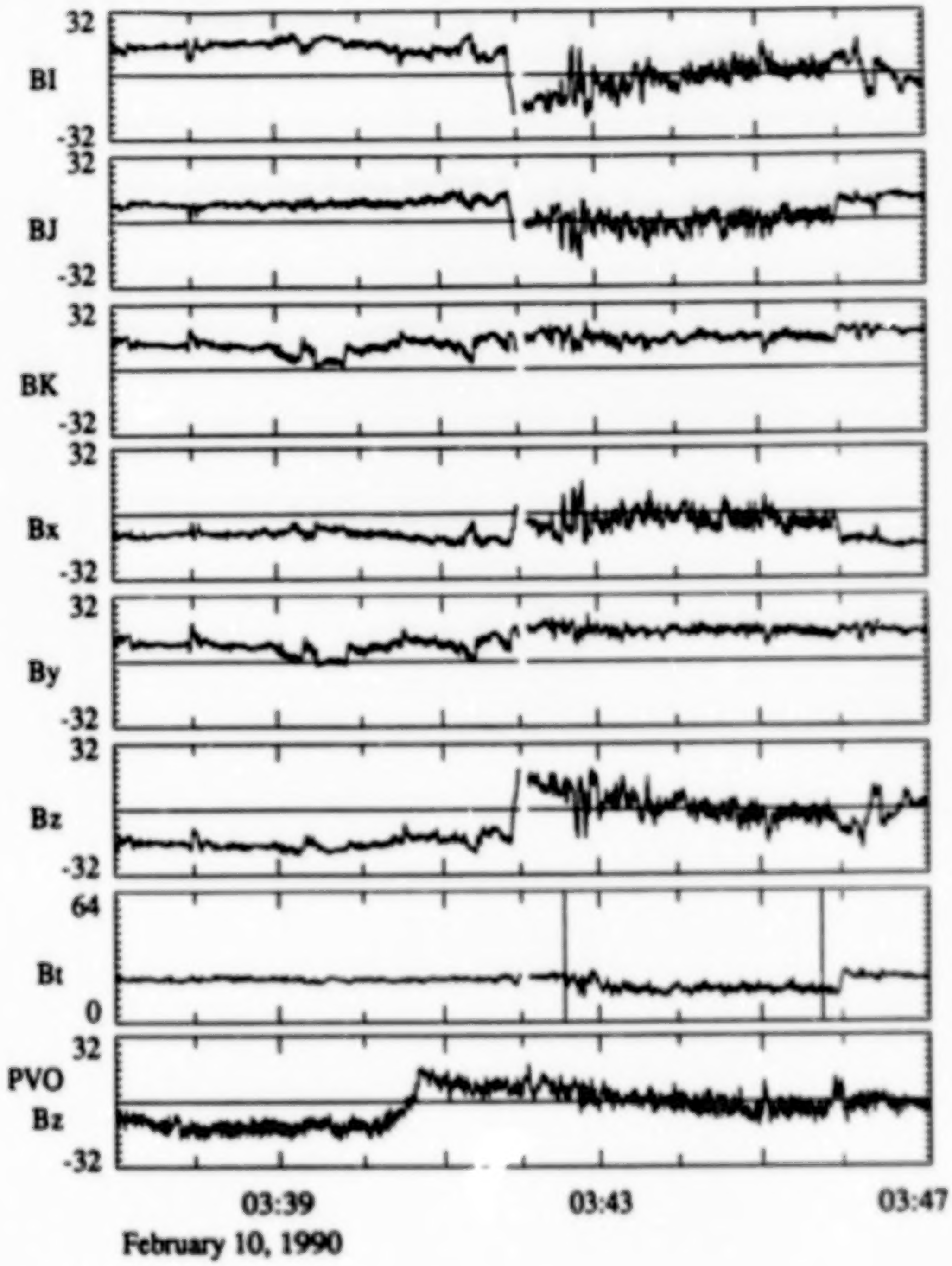


Figure 3. Magnetic field data in shock normal and VSO coordinate systems for the interval between 03:37 and 03:47 UT on February 10, 1990 [4]. The bottom panel shows the VSO B_z component observed by PVO. The interval used in the shock normal calculation is denoted by vertical lines on the B_T panel.

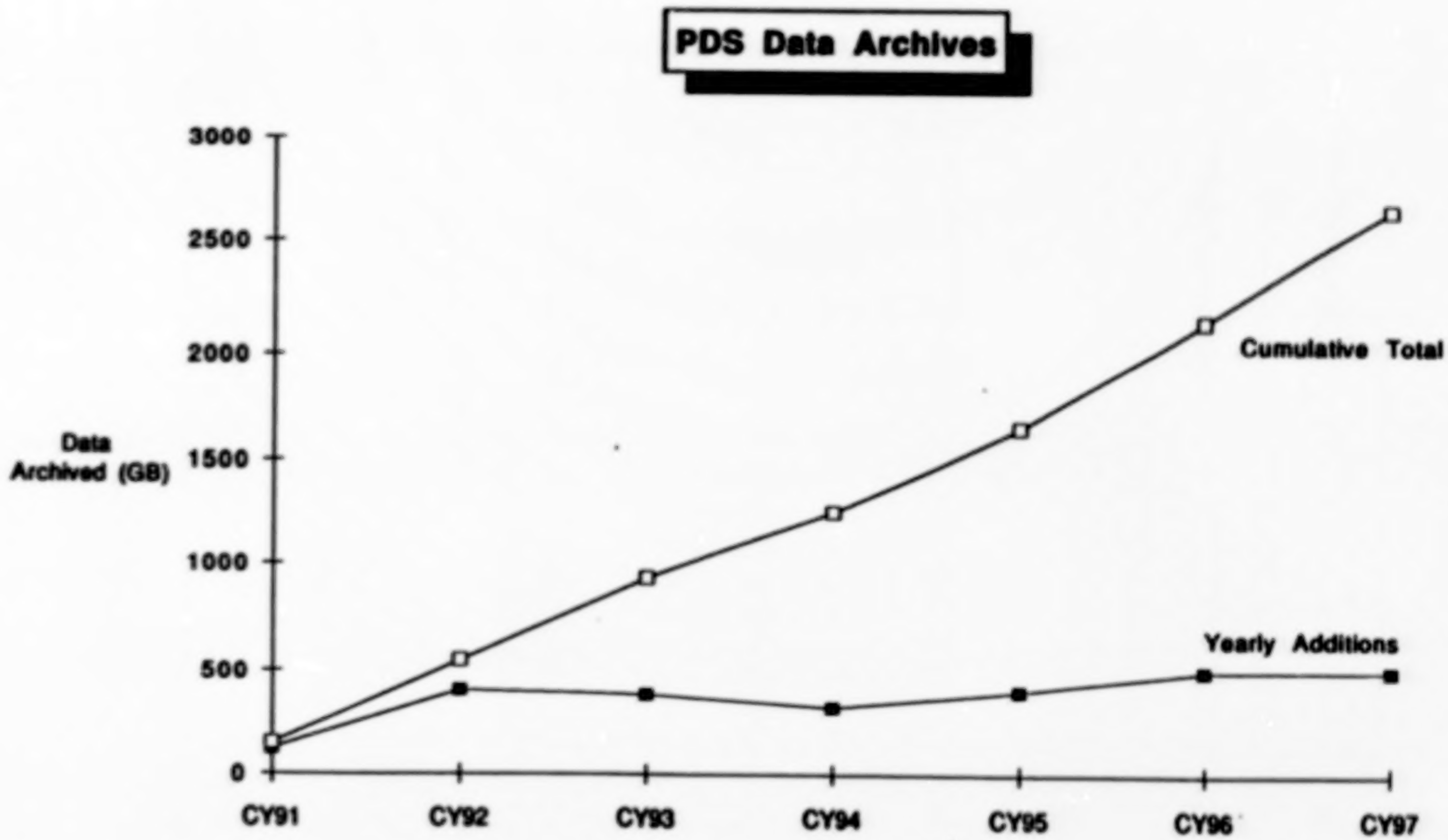


Figure 4. A projection of the data volumes to be archived by the Planetary Data System from 1991 to 1997 (courtesy of S. McMahon). The open symbols give the cumulative total while the solid symbols give the yearly additions.

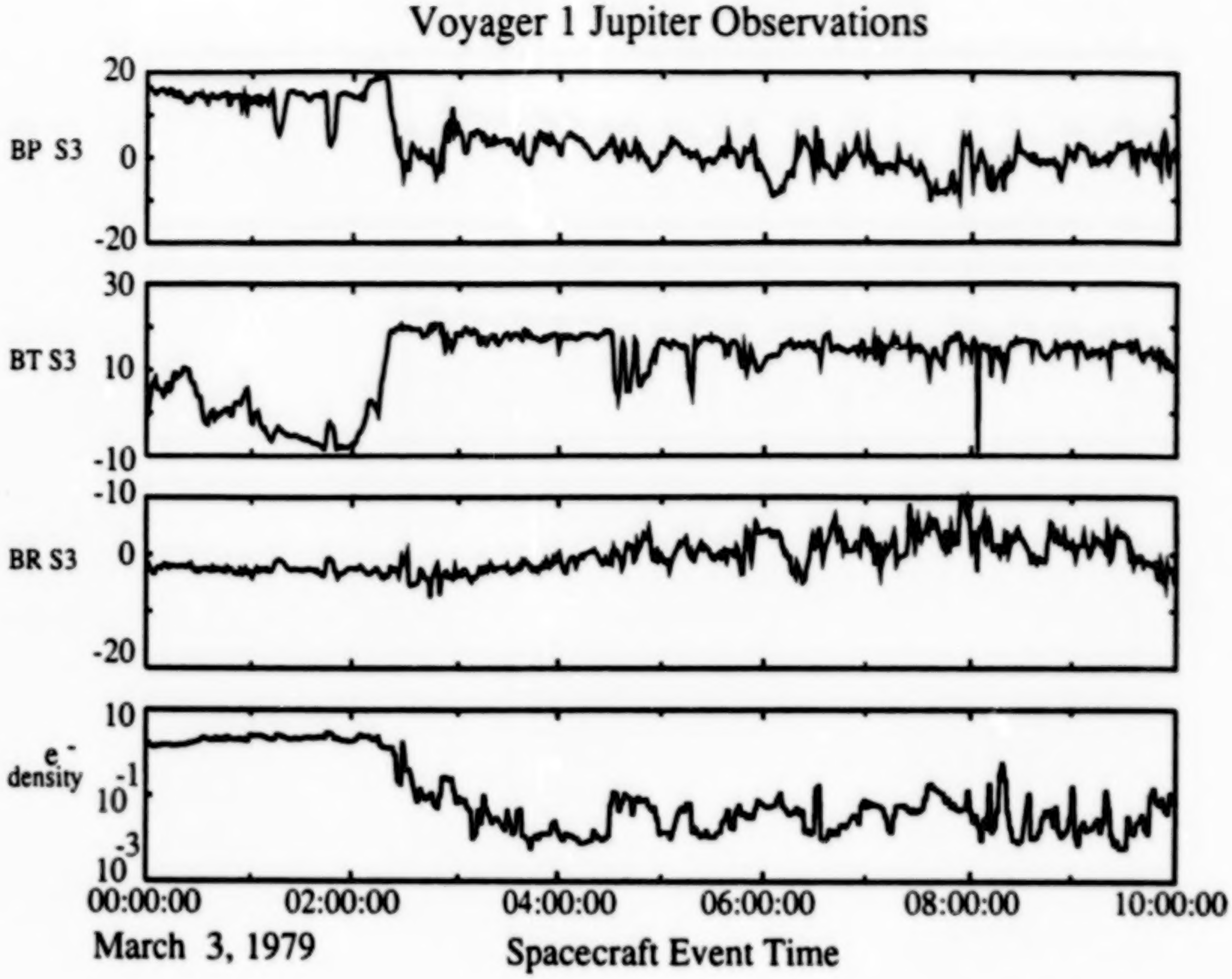


Figure 5. A typical data display from the Planetary Plasma Interactions Node Browse System. Plotted are magnetic field data in Minus System III coordinates and the electron density from the Voyager 1 encounter with Jupiter.

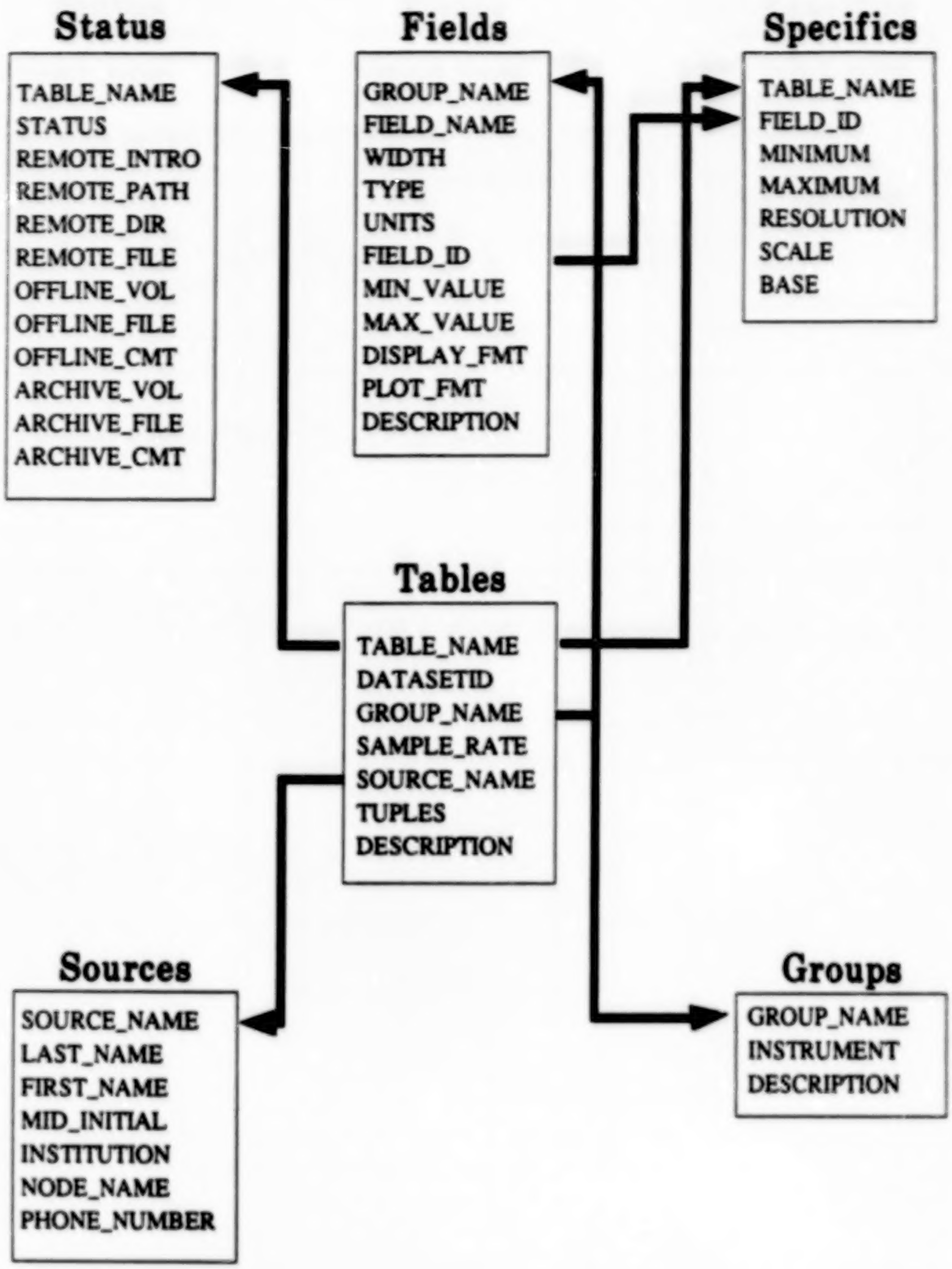


Figure 6. The file management tables used by the Planetary Plasma Interactions Node of the Planetary Data System. There are six tables (Tables, Fields, Status, Specifics, Sources, and Groups). The Tables table contains one entry for each table (data file) in the system. The Fields table contains the description for each field in a data table record. It is linked to the Tables table by the group_name field. Status contains data about the status of individual data tables controlled by the system. This includes the location of the data and whether it is on-line or off-line. The Specifics table contains information which is unique to each data table. It contains one entry for every field in every data table. The Sources table contains information about the source of the data contained in the table such as the name of the data supplier. The Groups table contains information related to data set groups. It includes a description of how the data were grouped (i.e., by spacecraft, target, etc.).

N92
12430
UNCLAS

MICROGRAVITY SCIENCE REQUIREMENTS AND THE NEED FOR DATA COMPRESSION

William G. Hartz
Analex Corporation
NASA Lewis Research Center
Cleveland, Ohio 44135

Abstract. The Microgravity Science and Applications Division (MSAD) of the NASA Office of Space Science and Applications (OSSA) is responsible for encouraging and directing the research of a wide range of physical phenomena in reduced gravity. Under MSAD's direction the NASA Lewis Research Center is presently developing the concept of a multi-user facility which will perform combustion science experiments in space. This facility, known as the Combustion Experiments Module (CEM), will be located in either the Shuttle Spacelab or the Space Station *Freedom* laboratory and will be operational by mid-1997. CEM shall be used to investigate the behavior of a wide range of combustion processes in the microgravity environment which exists in near Earth orbit.

In addition to standard instrumentation to measure temperature, pressure and acceleration, CEM shall employ a variety of imaging and optical diagnostic techniques. Images shall be the primary source of experimental data. Some preliminary experiment requirements indicate the facility may require up to five electronic cameras simultaneously generating images at 30 frames per second. Typically, each image will consist of 512 pixels by 480 pixels with 8 - 12 bits per pixel. In most cases the maximum experiment duration is on the order of 2 minutes. However, one experiment, investigating smoldering combustion, shall last up to one hour.

These images create an enormous amount of data which must be archived on orbit for later analysis. Additionally, ground based investigators will require enough data from the orbiting facility to determine if the experimental parameters need modification before proceeding with the next run. The storage and transmission of this data present a major challenge to the CEM design. Data compression will play an important role in the design of the CEM diagnostics system.

1. Introduction

This paper discusses the science data requirements for the Combustion Experiments Module. This set of requirements serves as an example of data required for microgravity experiments to be conducted upon either the Spacelab or Space Station *Freedom* in near-Earth orbit. Microgravity science research depends increasingly on full-field data which is captured in images. This is particularly true of the diagnostics proposed for combustion science research. In addition to images; instrumentation measurements, such as temperatures, pressures, and accelerations, must be recorded. Scientists require the entire data set to be recorded in the module on-orbit, and they also desire to have the entire data set downlinked. Still, the downlinked data should accommodate at least a "Quick Look" as a subset of the data between experiment runs. This capability is part of a concept known as telescience. In this concept, the principal investigator can interact with the experiment from a ground facility. The investigator will observe the experiment and its data, and he can communicate with mission specialists to modify experiment parameters.

The module will generate a great amount data. Also, the operation of the module, including telescience, will increase the downlink data rate. Limitations in data storage and in downlink capacity suggest a need for both lossless (for recording) and lossy (for downlinking) data compression. This paper identifies critical image and data parameters which must be maintained when considering lossy compression.

NASA's Office of Space Science and Applications (OSSA) funds research by university, industry, and government investigators in ground-based and space-flight facilities. This includes basic research in physical, chemical, and biological processes in a reduced-gravity environment. Investigators also perform basic and applied research on fluid dynamics, transport phenomena, and the processing of many materials and substances. OSSA's Microgravity Science and Applications Division (MSAD) develops space-flight payloads for the Shuttle, Spacelab and Space Station *Freedom*. Currently, payloads are developed to address the science requirements of a single investigator's experiment. However, MSAD's new focus is to develop payloads which are configured (or reconfigured) to accommodate multiple experiments. The Combustion Experiments Module is an example of this new payload. This move towards "laboratory" facilities occurs as increasingly sophisticated and complex diagnostics are being developed. Both of these lend to increased science data.

These orbiting facilities will permit investigators to conduct experiments in reduced gravity for long periods of time. This time period ranges for one minute to approximately one hour for proposed experiments in the Combustion Experiments Module. Current research in drop towers and aircraft limits experiments to two to ten seconds of microgravity. Also, the quality and level of the reduced gravity varies, somewhat unpredictably; especially in the case of aircraft experiments. Still, these platforms provide much information which leads to research conducted in near-Earth orbit.

MSAD payloads permit investigators to study physical phenomena, without buoyant flows, which can modify, mask or dominate a phenomenon in Earth's gravity. Investigators can also study and compare phenomena in Earth's and reduced gravity. Experimental data gathered in these payloads aid the development and verification of practical mathematical models. Some of the proposed MSAD modules are the Microscope, Containerless Processing, the Glovebox for small, self-contained packages, the Combustion Experiments Module, the Furnace Facility, the X-ray System, the Advanced Fluids Modules, Bio-technology, and Advanced Protein Crystal Growth. These modules cover a wide area of science and have a broad range of image and data requirements. The requirements of each module is unique to its science; yet, they have many similarities across disciplines.

2. The Modules and Why: An Example is the Combustion Experiments Module

As an example, this discussion focusses on the Combustion Experiments Module (CEM). The science data requirements for this module highlight a paradox: facilities in near-Earth orbit give longer time periods for the experiment and improved diagnostic capabilities yield large amounts of data; however, carriers such as Shuttle Spacelab and Space Station *Freedom* have limited downlink capacity and a data storage problem.

The Combustion Experiments Module (CEM) is a multi-user, modular facility which will accommodate several different experiments, each having numerous runs, during one Spacelab mission or one Space Station utilization cycle. Experiment hardware can be changed out during a mission. After a mission, the module can be reconfigured to run another set of experiments.

The design of the facility will also permit the on orbit changing of diagnostic instruments, optics, and cameras. Extensive diagnostic capabilities provide mapping of temperatures, velocities, and species concentrations. Many of these mappings result from images. Consequently, the CEM is a heavy user of video images, and it relies heavily on the accurate recording and interpretation of these images.

Combustion science differs from other branches of fluid physics because of large temperature variations, 300K to 3000K. Highly localized, highly exothermic heat release from the chemical reactions of the combustion process creates large temperature variations and large density gradients. These potentially lead to the strong currents of buoyant flows. The flows can dominate, modify, or mask the convective transport processes which mix and heat the fuel and oxidant reactants before chemical reactions begin.

Because of this complexity, buoyancy is often omitted from the mathematical analysis of combustion. Complicated two phase flows and surface tension behaviors are also affected by the buoyant flows. Gravity also introduces a degree of asymmetry in an otherwise symmetric phenomena. For example, combustion of a gaseous jet injected normal to the gravity vector quickly loses its axial symmetry as the flame plume gradually tilts "upward". Transport phenomena, feeding the flame, are multidimensional and complex.

The science data requirements for the CEM are provided as an example of the type of data needed in a microgravity experiment. It is also an example of the modular, multi-user facilities being developed by MSAD. While some of the particulars of an experiment or class of science may vary in number, types of measurements, storage, and data rates, the general scope of the experiments is similar.

3. A Summary of Science Data Requirements for the CEM

The particular data requirements discussed here are taken from the seven proposed experiments currently under consideration. Also, the particular diagnostic methods described here may vary as their development and the need for them continues.

CEM scientific data will come from two main sources: instruments and images. Table 1 shows a list of proposed experiments for CEM and the types of diagnostics each might use. The techniques are optical methods currently under investigation for space-flight. Specific diagnostic methods, as well as the experiments, may change by the time of the first launch in 1997. This table also shows an estimate of the type and number of instruments required for each experiment.

Table 1: Video and Instrumentation Requirements for CEM

EXPERIMENT	TECHNIQUE (Optical): Video (usually 2 views)	IR or UV Detection	Particle/Light Sheet Visualization	Rainbow Schlieren	Light Extinction/Soot Number	INSTRUMENTS (Quantity): Temperature	Pressure	Acceleration
Effects of Buoyancy on Laminar Gas Jet Diffusion Flames	•	•	•	•	•	9	1	3
Fundamental Study of Smoldering Combustion Spread	•	•				8	1	3
Diffusive and Radiative Transport in Fires Experiment	•	•				8	1	3
Studies of Premixed Laminar Flames	•	•	•	•		6	1	3
Ignition and Flame Spread of Liquid Fuel Pools	•	•	•	•		10	1	3
Droplet Combustion Experiment	•	•				1	1	3
Laminar Jet Diffusion Flame	•	•				?	?	?

The operating scenario for CEM calls for data archiving, quick-look, and downlinking capabilities. The investigators require the entire set of scientific data be recorded. For a quick-look, a portion of the data will be downlinked between experiment runs. Investigators can verify the success of experiments and they can see if they are achieving the expected results. Also, even with the best of models and analysis, investigators are often surprised to observe unexpected phenomena during a test run.

With a run length of approximately one minute, most experiments will end before the data subset can be viewed on the ground. Still, a quick-look will enable investigators to vary test parameters before the next run, thus maximizing the science return from an experiment.

The entire set of stored data may be downlinked at a later time when channel capacity is available. This is determined by the duration of the mission and the requirements of other payloads aboard the carrier. Also, data set downlinking will free up storage resources for subsequent CEM experiments. Finally, this capability may be used to guard against the loss of data. After the mission, "hard copies" of the data will be recovered including data storage media, film, videotapes, and experiment samples.

Estimates of the data rate and data storage needed for one CEM experiment challenge the limits of available storage capacities. They also greatly exceed the limits of available downlink capacity. These estimates do not include formatting, data tagging, or other types of headers or annotation which may be required.

The data rate for instrumentation (temperatures, pressures, flows, accelerations) range from 1.0 to 27.2 kilobits per second (Kbps). For images, the data rate varies from 670.0 Mbps (megabits per second) for 3 cameras to 1.0 Gbps (gigabits per second) for 5 cameras. The available downlink data rate aboard the Shuttle Spacelab varies between 1.5 to 48 Mbps. The expected downlink data rate for a Space Station payload is 48 Mbps. Some form of data compression will be necessary to achieve real time or near-real time transmission of on-orbit scientific data.

Estimates of CEM data storage requirements vary from 40 to 67 Gb (gigabits) per experiment. Data storage options, for either carrier, are analog tape and digital storage, possibly to 1.0 terabits. For monochrome images, a Super VHS video cassette (analog tape) will provide sufficient resolution and signal-to-noise ratio for many applications. In optical diagnostic methods where 24 bit, true color images are required, this type of analog recording may not be adequate. These images may require storage as digital images. The effects on data fidelity of recording these color images on a Super VHS machine requires additional study.

4. A Need for Data Compression while Preserving Data Fidelity

Use of sophisticated diagnostics, like the ones listed in Table 1, generate a large number of images in addition to more conventional instrumentation like thermocouples, pressure transducers, accelerometers, etc. This reflects the investigator's desire for field type measurements as well as point measurements. The science requires the correlation and annotation of data from these varied sources.

Lossless compression is preferred, and in some cases required, for data storage. Still, some lossy compression might be considered if storage capacity and data rates dictate the need for it. Greater compression, which is required for the Quick-Look in telescience, requires full motion in order to observe the phenomena of the experiment. In this case, the unexpected must be captured, so a technique which severely compresses the inter-frame motion is undesirable.

However, more central than the question of lossless or lossy compression is the question of the impact compression has on the fidelity of the data derived from the images. As suggested earlier, some signal reduction or degradation may have little impact on the accuracy of the final data analysis. This entire question requires further investigation. Most importantly, implementations of compression must

preserve critical characteristics of the image. This may be to increase the intensity of a weak signal; for example, so that a dim flame can be distinguished from its background. When determining soot number, small changes in the intensity of a light beam passing through a flame are directly related to the amount of soot produced in the reaction. In many experiments' images, the high spatial frequency information associated with edges needs to be preserved. In true color, 24-bit images, the transformation of color coordinates can permit significant compression. In all cases, images will benefit from processing and compression methods which minimize noise and increase or preserve the dynamic range. They require robust algorithms which protect images from over-optimization. This is true, especially when the images will be further processed in the downlink. It is possible to compress an image to the point where a transmission error is fatal to the transmission of an entire image or series of images.

In combustion science, one class of experiments investigates phenomena resulting in very faint, low luminosity flames. Compression or processing would be desired which can help extract this faint signal from the background. This increased intensity or dynamic range should be done without increasing the noise level of the recording medium. This is the current problem when motion picture film is push processed to reveal faint flames. Also in slow flows, techniques are needed which help to discriminate particles over time. Quite often, these small particles, which are used to indicate the velocity of flows, become difficult to distinguish from the "background" of the fluid in which they are moving.

Techniques which help to preserve and discriminate edges are useful in the analysis of flame front propagation. This is a critical parameter in the study of solid smoldering and combustion. These techniques will also help measure the change in diameter in burning fuel droplets.

Color is another important characteristic of images to preserve. Natural, high fidelity color indicates much about the type of combustion and chemical reactions occurring. The accurate determination of hue with reference to a calibrated reference is critical for diagnostic methods such as rainbow schlieren deflectometry. In this technique, the RGB (red-green-blue) camera signal is converted to an HSI (hue-saturation-intensity) signal. The hue is of primary importance. In this case, a 24-bit, true color, RGB image can be significantly reduced to an 8-bit hue signal with fewer than 8 bits each for the saturation and the intensity signals.

The rainbow schlieren technique indicates how preprocessing of data might be used to reduce the number of bits per pixel. The hue in the resulting image is compared to a reference image. The corresponding hue indicates the amount of deflection of the light beam due to changes in the index of refraction along the beam's path. This information is used to determine temperature and density variations across the field of the light beam. However, this type of preprocessing is unique to a particular technique and is different from a general compression strategy. The hardware associated with this type of technique and its preprocessing is inserted in the signal stream before data storage or downlinking. It affects the initial data; and therefore, pre-processing on orbit may not be desired.

Compression algorithms require implementations which are efficient with short processing times. The limitations of volume and power in orbiting facilities such as CEM indicate the need for routines which efficiently run with minimum memory and low electrical power. Hardware implementations, although lacking in flexibility and requiring development time, provide speed and efficiency in a low power package. Often the memory is integrated with other processing elements into the package.

5. Conclusions

The science requirements for CEM serve as an example of the data required for MSAD's near-Earth, reduced gravity experiments. Although the particulars of an area of scientific research may differ, the general data handling problems are the same across modules. Often as scientific research progresses, diagnostic methods become more sophisticated with more steps in the analysis of the data. In many

instances, these techniques involve video images. All of the modules have requirements for image recording and transmission. The increasing use of electronic cameras and image analysis amplifies this problem.

This discussion indicates the need for further research into the applications of image compression and processing within the orbiting module. One issue is to assess the impact of recording, compression, and processing on the fidelity of data for full field measurements. Likewise, one needs a means to assess the effects of these processes on image quality. Compression and processing of images must preserve important features such as edges, color and intensity. These processes need to preserve dynamic range and to maintain or increase signal to noise ratios. These factors will help to preserve image quality and data fidelity.

The concept of telescience, which enables the scientist to observe and conduct the experiment from the ground, will require some type of data compression for the downlink. Experiment automation and telescience also make increasing use of electronic imaging and image analysis. Future experimental scenarios, and the weight and volume constraints on the amount of film or videotape carried to spaceflight, increase the need for downlinking and recording of data off of the carrier. Data compression, if properly applied, can provide a solution to the data storage and transmission problems of on-orbit experiment facilities such as CEM.

DISCUSSION GROUP REPORTS

omit

THE DISCUSSION GROUPS

1. Organization

Before the workshop, the workshop organizers suggested six discussion topics for the group discussions. These were:

1. Data Compression for Browse/Quick Look,
2. Data Compression for Data Archival,
3. Data Compression as a Pre-Analysis of Space and Earth Science Data,
4. Data Compression for Near Earth to Earth Transmission,
5. Data Compression for Deep Space to Earth Transmission, and
6. Techniques for Containing Error Propagation in Compression/ Decompression Schemes.

As the participants registered for the workshop they were asked to indicate their first and second choices for their discussion group topic (space was given for indicating an alternative topic, but no one indicated such a topic). According to the interest indicated by the participants, topics 2 and 3 were combined before the workshop, as were topics 4 and 5. Further, when the discussion groups were actually formed at the workshop, topics 1 and 2 were combined, and topic 3 was dropped.

The final discussion groups and group leaders were:

1. Data Compression for Data Archival and Browse/Quick Look, Jeff Dozier and James C. Tilton.
2. Data Compression for Near Earth and Deep Space to Earth Transmission, Daniel E. Erickson.
3. Techniques for Containing Error Propagation in Compression/Decompression Schemes, Ben Kobler.

2. Goals

The first goal of each discussion group was to examine the potential for data compression to address data storage and transmission constraints found throughout the domain of NASA missions. The second goal was to recommend specific actions directed at enabling mission use of appropriate data compression technologies to overcome these constraints.

3. Participants

Each group comprised a nearly equal mix of technologists and users. The data compression technologists provided expertise in the current state of the art of the technology. The users, mostly designers of data systems and spaceborne experiments, provided an understanding of the broader issues of requirements, system constraints, and future requirements trends. The participants came from NASA, universities, and industry. The names of participants in each discussion group are given at the end of each discussion group report. The appendix lists the names and addresses of all participants in the workshop.

4. The Discussion Process

Each group began its considerations by identifying key technical issues which either could be addressed by data compression or inhibited the incorporation of data compression on NASA missions. It then proceeded to list actions and programs which would support the evaluation, development, and use of data compression technologies. After identifying which issues were addressed by each action, the group recommended a small set of actions and programs. Some of this work took place after the workshop, in

the process of reviewing this summary. For the sake of brevity, only those issues and actions which the group feels would have the greatest overall effect are discussed here. However, since data compression is very application dependent, there are so many examples that every case cannot be covered in a brief report. Lack of mention in this summary does not constitute an anti-endorsement. This application dependence also means that often a modest investment in a niche application can have dramatic results.

N92

12431

UNCLAS

S6-82
N92-12431
568
5P

DATA COMPRESSION FOR DATA ARCHIVAL, BROWSE OR QUICK-LOOK

Jeff Dozier
Universities Space Research Association
Goddard Space Flight Center
Greenbelt, MD 20771

James C. Tilton
Goddard Space Flight Center
Greenbelt, MD 20771

1. The Applications

1.1 Archival

Soon after space and Earth science data is collected, it is stored in one or more archival facilities for later retrieval and analysis. Since the purpose of the archival process is to keep an accurate and complete record of data, any data compression used in an archival system must be lossless, and protect against propagation of error in the storage media. In contrast, browse and quick-look require only the retrieval of a good approximation of the data, allowing consideration of lossy data compression. What is a good approximation depends, of course, on the data characteristics and the purposes for which the data is being browsed or previewed.

1.2 Browse

A browse capability for space and Earth science data is needed to enable scientists to check the appropriateness and quality of particular data sets before obtaining the full data set(s) for detailed analysis. Browse data produced for these purposes could be used to facilitate the retrieval of data from an archival facility. Appropriately derived browse data can also facilitate interdisciplinary surveys which search for evidence of unusual events in several data sets from one or more sensor. Such browse data can also be used to validate the quality of the data by facilitating quick checks for data anomalies.

1.3 Quick-look

Quick-look data is data obtained directly from the sensor for either previewing the data or for an application that requires very timely analysis of the space or Earth science data. This quick-look data could be either a small subsection of the full resolution data, or an approximate representation of a larger section of data, such as described for browse data. In the latter case, lossy data compression techniques tailored to retain the information significant to the particular application would be appropriate. Two main differences between data compression techniques appropriate browse and quick-look cases are the quick-look techniques (i) can be more specifically tailored, and (ii) must be limited in complexity by the relatively limited computational power available on space platforms.

2. Key Issues

2.1 Archival

Storage space: If lossless encoding is required, possible compression savings are limited to approximately 2:1 for most space and Earth science data. If this is the only justification for data compression, the use of data compression may not be justified since one could just buy twice as much of the storage media.

Data integrity: Any encoding of the data must be robust to errors in the storage media, and must retain the full scientific information content of the original data. For experimental data, this would generally mean that every bit of the original data must be retained.

Data access: Quick access is required to information about archived data, allowing interactive ordering of data from the archive. Appropriate browse data product(s) could serve to augment other descriptive data that is kept on-line for fast access, while the full data set is kept in off-line storage. Algorithms for decoding the compressed browse or full resolution data must be very fast. However, encoding speed is not critical, since there will be many decodes per encode.

Synergism: If decrease in storage space does not justify the use of data compression, a system employing data compression as an integral part that decreases storage space requirements, increases data integrity and improves data access would most certainly be justifiable.

2.2 Browse

Facilitate Access to Archived Data: Essential information for a wide variety of applications must be retained in the browse data for widest utility. A multitude of scientific data products may be generated from most space and Earth Science data sets. In addition, space and Earth Science data sets come in several different forms, including images, time series, 3 or 4-dimensional data, and housekeeping or ancillary data. For efficiency, browse data compression must be well integrated into the archival/data access facility. A well integrated browse facility would enable interactive ordering of archived data, and speed access over remote networks. In such a facility required information could be retained on-line for quick access.

Search for Unusual Events or Data Anomalies: Browse data produced by approaches that smooth the data too much, or bias towards expected or previously observed data signals, are not acceptable for these purposes.

Browse Data Quality: What quality is required? Can scientific analysis be performed on browse data? Can the production of browse data be made sufficiently "smart" to retain the information required for at least a preliminary scientific analysis of the data? The effects of the lossy compression used to produce the browse data must be analyzed for the effects on the results of the scientific analysis of the data (rather than just visual appearance).

Modes of Access: The user may want to be able to compare visually many browse images at one time, and then select one or more for more detailed analysis. Alternatively, the user may want to look at large portion of a data set in browse mode, and then focus done to a smaller subset for more detailed analysis.

2.3 Quick-Look

Computational Complexity: Quick-look can most easily be done as a rapid transmission at full resolution of a small subset of the data. When doing more than subsetting the data, the encoding algorithm must be limited in complexity by the relatively limited computational power available on space platforms. It is difficult to space qualify more powerful computer hardware.

Tailoring: Since quick-look data would be used for a specific purpose, the production techniques can be specifically tailored to the application.

2.4 Other

To facilitate wide participation in the development process, NASA data compression systems should follow accepted standards as closely as possible, such as JPEG (Joint Photographic Experts Group) or MPEG (Moving Picture Experts Group).

3. Data Compression Approaches

3.1 General Approaches

The data compression field is already highly developed. Given here, instead of a review of techniques, is a bibliography books on compression recommended provided by Robert M. Gray:

Lossless Data Compression (Noiseless Coding):

J. Storer, *Data Compression: Methods and Theory*, Computer Science Press, 1988.
T. J. Lynch, *Data Compression: Techniques and Applications*, Lifetime Learning, Wadsworth, 1985.

Transform and Predictive Coding:

N. S. Jayant, ed., *Waveform Quantization and Coding*, IEEE Press, 1976.
N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.
R. J. Clarke, *Transform Coding of Images*, Academic Press, 1985.
A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression*, Plenum Press, 1988.

Vector Quantization:

H. Abut, ed., *Vector Quantization*, IEEE Press, 1990.
M. Rabbani and P. Jones, *Digital Image Compression*, SPIE Publications, 1991.
A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer, 1991.

3.2 Progressive Transmission

Progressive transmission techniques are a natural match to efficiently combining browse and data archival. Progressive transmission techniques can losslessly encode data, but the early stages of reconstruction naturally produce choices of data renditions that could be used as a browse version of the data. If none of the renditions is satisfactory as the browse version of the data, other means could be used to produce the browse version, and the difference between the browse data and original data could be losslessly compressed by progressive or other means. In either case only the information required to produce the browse rendition would be kept on-line, while the remainder of the information required to reproduce the original data would be retained in off-line storage.

3.3 Synergism with Analog to Digital (A-D) Conversion

Nearly all Space and Earth Science data collection involves A-D conversion. Since A-D conversion is in itself a gross form of lossy data compression, gains in information content per volume of data may be obtained by combining more sophisticated forms of lossy data compression with A-D conversion. The current approach using a uniform (or perhaps companded) quantizer for A-D conversion followed by lossless compression (if compression is employed) is suboptimal. An example of employing lossy compression techniques to optimize this process would be convert the analog signal into vector codes, such as done in vector quantization (a form of lossy compression). Vector quantization design techniques could then be employed to tailor the overall source code to characteristic of the data being encoded.

3.4 Other

If a large amount of on-board memory is available, a possible approach to data compression would be to just transmit the changes observed in the data from the same location from one orbit to the next.

Besides large amounts of on-board memory, this approach would require sufficient computational power to register the data collected in the current orbit with that from the previous orbit.

4. Open Questions

How predictable is a time series of images when the time interval is days, rather than seconds or split seconds? Can we losslessly compress a time series of, for example, MODIS data?

How can a browse system be designed intelligently so various types of remote sensing data (SAR data, multi-spectral data, or spectrometer data), time series data (with small time intervals), or housekeeping/ancillary data are handled appropriately?

5. Recommendations

There is a critical need to promote interaction between data compression scientists and space and Earth scientists to more effectively explore the utility of data compression techniques for space and Earth science data. A first step that can be done immediately (without specific new funding) is for NASA to provide test data sets and examples of analysis scenarios to data compression scientists. This data and scenario information could be kept at an "anonymous ftp" site, and/or made available on an optical disk. At a minimum, this will enable researchers to determine if their existing techniques are, or are not, appropriate for space and Earth science data. A more structured (i. e., funded) program would be required to insure feedback and more intensive refinement of approaches to suit the data and analysis scenarios. Possibly this effort could tap into the Version 0 EOSDIS activity. An important task to be accomplished by a more structured activity would be to statistically characterize the various classes of space and Earth data.

Certain technical approaches stand out as being particularly promising. The application of data compression to browse and data archival is one. Development of this type of system for various data types should be promoted. Also to be encouraged is the production of "smart" browse data for various different data types and applications. This "smart" browse data would retain most of the essential information for a rough, but still informative, scientific analysis of the data. This research would provide feedback concerning the best types of browse data to provide as an integral part of a data archival access system.

Another area of research that should be encouraged is the combination of lossy compression techniques with analog to digital (A-D) conversion.

We recommend that NASA should make the pursuit of research in these and other promising areas related to the compression of space and Earth science data an area of emphasis in one or more future solicitations (e.g., NASA Research Announcement) under the Applied Information Systems Research Program and/or other appropriate NASA program.

The organizers of the Data Compression Conference, of which this workshop is a part, have already announced that the next Data Compression Conference (DCC'92) will be held on March 24-27, 1992 in Snowbird, Utah. We recommend that participants in DCC'92 be encouraged to test their methods on a standard set of images provided by NASA. This standard set of images might include Landsat Thematic Mapper images, AVIRIS images, SAR images, space time series data. Perhaps some "bad" data should also be included. A special session at DCC'92 could be devoted to discussing and contrasting these results.

Participants

The participants in this discussion group were Karen Anderson, Joe Bredekamp, Mayun Chang, Pamela C. Cosman, Linda Jo Dolny, Jeff Dozier, Benjamin Epstein, Wai-Chi Fang, Robert G. Finch, Richard Frost, Daniel Glover, Robert Gray, Paul G. Howard, Peter Kenny, Manohar Mareboyana, Kristo Miettinen, Jeff Niehaus, Karen Oehler, Dale Rickman, Eve Riskin, S. Srikanth, Vincent Salomonson, Rick Schumeyer, James C. Tilton, Jeff Vitter, and Ray Walker. See the appendix for addresses.

N92
12432
UNCLAS

57-82
51319
N92-12432
p-6

DATA COMPRESSION FOR NEAR EARTH AND DEEP SPACE TO EARTH TRANSMISSION

Daniel E. Erickson
Jet Propulsion Laboratory
Pasadena, CA 91109

1. The Applications

1.1 Near Earth Satellites

Communications Capabilities: In the foreseeable future, near Earth polar and equatorial satellites will communicate to the ground via the Telemetry and Data Relay Satellite System and its successors. TDRS can support up to 300 megabits per second of dedicated transmission. Contention for this high-rate communication resource will limit access by any one satellite. The TDRS also has several lower rate channels which can allow access by multiple satellites. Data may also be dumped at high rate to Ground Tracking and Data Relay Stations as the satellite passes through their range. Some satellites may also support direct downlink of timely local data to small ground stations. Direct downlink transmission will be at data rates of only a few megabits per second, to allow small inexpensive receiving stations.

Communications Drivers: Several instruments which have been considered for Earth Observation have high raw data rates. The Synthetic Aperture Radar (SAR) instrument takes data at over 300 megabits per second. The High Resolution Imaging Spectrometer (HIRIS) instrument takes data at 420 megabits per second. Of additional concern are instruments with lower data rates but high data volumes because of high duty cycles. The Moderate Resolution Imaging Spectrometer (MODIS) instrument, for instance, takes data at 20 megabits per second continuously. Uncompressed, the MODIS data would take 40% of the average Earth Observing System (EOS) platform total downlink volume. Near real time direct downlink data are desired for ice data for navigation purposes, for regional pollution, rainfall and crop data, and remote sensing data for field experiments.

1.2 Spacelab & Space Station Freedom

Communications Capabilities: The space station Freedom will communicate with Earth at 50 megabits per second.

Communications Drivers: Potentially, the most data intensive activities related to the space station will be remote operation of scientific experiments. In this operating mode, sometimes called *telepresence*, principal investigators on the ground observe the progress of space based experiments and direct them either through electronic commands or through voice communication with the astronauts. In order to direct the experiment, the P. I. needs information on the progress of the experiment, possibly through real time video. Full color video, uncompressed would take 46 megabits per second per video channel. Remote monitoring is desirable for microgravity and life sciences experiments. In addition, microgravity experiments may require non real time high resolution, high rate video to meet science objectives.

1.3 Geostationary Platforms

Communications Capabilities: Geostationary platforms would probably communicate directly with ground stations. They might even act as relays for satellites in low earth orbit. Several communications options could be available in the first decade of the twenty first century when the geostationary platforms are planned. Optical communications with spatial diversity to reduce the intervals of

PRECEDING PAGE BLANK NOT FILMED

blockage due to weather could achieve rates on the order of a gigabit per second. Near real time direct downlink to field sites would still have significantly constrained communication rates.

Communications Drivers: Geostationary Earth observation platforms will tend to have staring instruments with wide, continuous coverage. These will be based on EOS instruments and may be capable of very high data rates.

1.4 Lunar Base

Communications Capabilities: Bases on the near side of the moon will communicate directly to ground stations on Earth or with Earth orbiting relay satellites.

Communications Drivers: A lunar base would conduct experiments, explore the lunar surface, and make astronomical observations. The experiments and exploration would benefit from telepresence. The observations may have very high raw data rates.

1.5 Deep Space

Communications Capabilities: The data rates from interplanetary spacecraft are limited by spacecraft and ground based antenna size and constrained spacecraft transmission power. The highest data rate planned for the Galileo Spacecraft at Jupiter is 134 kilobits per second. Missions such as a Neptune orbiter face even lower data rates unless new technologies such as optical communications can be developed. With optical communications, rates on the order of a megabit per second can be hoped for.

Communications Drivers: Imaging has put the highest demand on downlink resources in recent missions. As we move to more detailed studies of the planets, moons, asteroids, and comets of our solar system, multispectral imaging and synthetic aperture radar, both data intensive instruments, will be desired.

2. Key Issues

2.1 Error Susceptibility

Data compression, even the lossless approach, increases the impact of bit errors in the communication link. This is due to the increased information content per bit. For some approaches, this effect is further exacerbated by the interdependence of the bits in the reconstruction of the data. By choosing the right approach and adding channel coding to the communication link, the net effect of compression and error coding can be better data quantity and quality at the cost of additional system complexity.

2.2 Data System Considerations

Some of the potential benefits of data compression can only be realized if the data system is designed to exploit them. Lossless compression, for example, produces a variable volume output. To fully exploit the reduction in bits required to send the desired information, the data system would need to handle variable length packets and prioritized telemetry.

2.3 Operations Complexity

The capability to use data compression expands the trade space which can be considered during operations. While the additional capability may ease some operation problems, the additional decision complexity may add a burden.

2.4 Quality versus Quantity Tradeoff

Lossy compression introduces the option of increasing the volume of information which can be sent to the ground at the cost of adding distortion. NASA scientists cannot yet assess the impact of such trade-offs. Furthermore, this assessment is application dependent. Several lossy data compression schemes have been studied in academia and industry. One fact which has become clear is that the performance of a compression algorithm, in terms of reduction ratio versus quality, depends on the statistics of the data being compressed and the quality function appropriate to the application. Some schemes preserve edges and fine scale features, for instance, where others blur them or treat them as noise. Which approach is more satisfactory depends on the use to which the data will be put. While a distortion measure such as root mean square error is statistically precise, it is not always the appropriate measure of quality.

2.5 Experiment Design Considerations

Assuming that an instrument has been allocated a fixed bandwidth, designers are faced with several alternatives:

a) Design an instrument incapable of exceeding the allocation,
b) Design a more capable instrument, but use it only part of the time and provide rate buffering (duty cycling),
c) Delete spatial or spectral components or decrease precision (editing),
d) Accumulate data, lowering spatial, spectral or temporal resolution (integrating),
e) Compress data in a manner which allows exact reconstruction (lossless compression),
f) Compress data in a manner which introduces distortion (lossy compression),
g) Reduce the data by on-board parameter or feature extraction (data processing).

Probably, a combination of the above techniques will give the best performance for the cost. The capability to perform on-board data processing or lossy compression is just now becoming a reality. Scientists have not yet considered what experiments might be enabled by combining these options with more powerful instruments.

2.6 Spacecraft Resource Considerations

Mass, power and volume are often scarce resources in spaceborne systems. While a single chip solution to lossless compression has been demonstrated, most compression schemes are more complex. At low rates, much can be accomplished by software on general purpose processors. The largest payoff, however, would come from compressing high-rate data. Many compression schemes appear straightforward enough to be implemented in a single chip or a small number of chips. This would reduce their use of spacecraft resources to an acceptable level.

2.7 Cost/Risk

While the non-recurring development cost and the recurring costs of including data compression on spacecraft may appear to be a barrier to doing so, this may be largely illusory. The cost per bit of information returned is significantly less than for many communications enhancements which NASA has funded over the years. (See Table 1.) Furthermore, the cost risk of adding compression is no greater than that of adding other new technologies. The performance risk for adding lossless compression is very low. The effects of lossless compression on the value of the returned data is well understood. For lossy compression of science data, however, the effect is not well understood in most cases. Lossy compression of operational data such as real time video and voice is much better understood and is being used commercially.

Table 1 gives cost/performance estimates for a number of improvements to the Deep Space Network (DSN). The unit of performance is a Big Aperture Performance Unit (BAPU), equivalent to one 70-meter antenna at 25 degrees Kelvin. Assuming that a 2:1 lossless compression were achieved, the effect would be equivalent to doubling the current capacity of 4.4 BAPUs. Experiments with data provided by the CRAF/Cassini and SIRTF projects and from the AVIRIS instrument have yielded lossless compression ratios ranging from 1.3:1 to 3.2:1, with 2:1 being a good conservative average if data are preconditioned to remove detector discrepancies.

Table 1. Performance versus Cost of Enhancement Techniques for the Deep Space Network*

TECHNIQUE	~BAPU gained	~COST $M	$M/BAPU
Upgrade all 3 64m to 70m	1.2	38	32.0
Array with VLA	2.0	20 (1st rental use)	10.0
Big Viterbi Decoder	1.6 (equiv.)	13	8.0
Compress all data 2:1	4.4 (equiv.)	5+3/mission	1.1
BWG and UNLAs on 70m	2.3	34+6	17.0
Ka-band and BWG on 70m	9.0 (equiv.)	27+10+5/mission	4.0

3. Solution Approaches

Several approaches to eliminating barriers to effective use of data compression were considered. The paragraphs below describe, not in priority order, those which the discussion group deemed most promising. Table 2 shows the issues which each approach would address.

3.1 Develop New Data Compression Techniques

While many data compression approaches are being explored commercially and in academia, NASA has several unique requirements which have not been fully addressed. High ratio compression would have a high payoff for remote experiment monitoring. Lossy compression which preserves science value could be important for a number of instruments, providing we could learn how to measure science value. Combining data compression (source coding) with error protection (channel coding) may yield more efficient use of communication and storage resources.

3.2 Improve Our Understanding of the Science Value of Compressed Data

Experimentally compressing realistic science data and determining the resultant effect on the analysis of these data would help to clarify and quantify the impact of proposed compression schemes. Studies examining the trade-offs involving more capable science instruments and observation/compression/analysis scenarios would help to clarify the alternatives for space and earth science observation.

---

* Data provided by Ivan Onyszchuk in memo 331-91.2-023 to Dan Erickson dated April 30, 1991

3.3 Develop Data System Designs and Operations Strategies for Data Compression

Variable length packets, optional prioritized telemetry, and event-responsive operations could allow the gains of on-board processing to be fully realized. Designs and demonstrations of such capabilities are needed to lower the risk of their incorporation into flight projects.

3.4 Develop Efficient Data Compression Hardware

To address the mass, power and volume constraints, data compression would best be implemented with application specific integrated circuits (ASICs). Flight qualifiable ASICs could be developed in the technology program for a few key compression techniques.

Table 2. Issues Versus Approaches for Compression Technology for Space to Earth Transmission

Approach / Issue	New Techniques	Science Value Studies	System Approaches	Compression Hardware
Error Susceptibility	X		X	
System Considerations			X	
Operations Complexity			X	
Quality vs. Quantity Tradeoffs	X	X	X	
Experiment Design Considerations		X	X	
Spacecraft Resource Constraints				X
Cost/Risk			X	X

4. Specific Recommendations

The discussion group on data compression for space to earth transmission makes the following recommendations:

1) Data compression is a cost-effective way to improve communications and storage capacity. NASA should use lossless data compression wherever possible. NASA should continue working with the Consultative Committee for Space Data Systems to define lossless data compression standards, so that space qualified hardware can make maximum use of commonality.

2) NASA should conduct experiments and studies on the value and effectiveness of lossy data compression. These studies should include participation by key earth and space scientists who would evaluate the decrease of science value due to the distortions introduced and the increase in science value due to increased temporal, spectral, spatial and measurement resolution and increased coverage. These studies might best be funded jointly by codes S and R.

3) NASA should develop and select approaches to high-ratio compression of operational data such as voice and video.

4) NASA should develop data compression integrated circuits for a few key approaches identified in the preceding recommendations.

5) NASA should examine new data compression approaches such as combining source and channel encoding, where high-payoff gaps are identified in currently available schemes.

6) Users and developers of data compression technologies should be in closer communications within NASA and with academia, industry, and other government agencies. A data compression working group, newsletter, and/or electronic bulletin board should be considered.

Participants

The participants in this discussion group were Daniel E. Erickson, William G. Hartz, Dana Kloza, Trent Mills, Dmitry A. Novik, Ivan Onyszchuk, Christopher J. Pestak, Robert Stack, Jack Venebrux, Wayne Whyte, Jr., and Carol Wong. See the appendix for addresses.

TECHNIQUES FOR CONTAINING ERROR PROPAGATION IN COMPRESSION/DECOMPRESSION SCHEMES

Ben Kobler
Goddard Space Flight Center
Greenbelt, MD 20771

1. The Application

The typical raw Bit Error Rate (BER) for space communications is one bit in $10x10^5$ bits. Through error correction of header information, this can be reduced to one bit in $10x10^8$ bits; through error correction of the whole data set this can be further reduced to one bit in $10x10^{12}$ bits. Similarly, the typical raw BER for archive media is one bit in $10x10^6$ bits; through error correction this can also be reduced to one bit in $10x10^{12}$ bits.

The total EOS data volume, however, is at minimum $10x10^{16}$ bits, to be accumulated over a 15 year period. If BER were to stay at one bit in $10x10^{12}$ bits, this would result in several uncorrectable errors per day. To avoid this, we must push toward better error correction. However, since we will also be doing data compression to minimize transmission and storage requirements, we have to understand the relationships between error correction and data compression.

2. Key Issue

Data compression has the potential for increasing the risk of data loss. Although data compression reduces the number of bits required for transmission and storage -- and hence the number of bit errors that can be expected -- data compression can also cause bit error propagation, resulting in catastrophic failures. For example, entire images could be rendered useless due to a single bit error. Techniques to detect these errors in compressed data and to minimize the resulting error propagation often involve trade-offs against compression performance.

3. Approaches

There are a number of approaches possible for containing error propagation due to data compression.

1) Data re-transmission - Requests for data re-transmission are only useful, however, when errors are detected, and only when errors are detected early. In space communication retransmission is often impossible; in archive systems re-transmission is often not helpful since the media may already be corrupted.

2) Data interpolation - Data interpolation is also only possible when errors can be detected. In addition, since we often have entire images destroyed, this may require data interpolation between entire time sequenced images -- a difficult technical task, and one that the science community would find difficult to accept.

3) Error containment - Error containment is already done to varying degrees in some data compression algorithms. Vector quantization, for example, sends compressed data in fixed sized blocks, thus limiting error propagation. Some Huffman codes allow quick error detection and re-synchronization, as does the DCT (Discrete Cosine Transform) JPEG (Joint Photographic Experts Group) algorithm which has an appended delimiter pattern and the Rice algorithm which has a fixed line format. Arithmetic codes, however, although efficient in compression performance, do not provide error containment. While this

can be improved via piecewise arithmetic coding, it is done at the expense of reduced compression performance.

4) Error correction - Error correction could be improved so the BER is perhaps only one bit in 10×10^{14} or 10×10^{15} bits. Errors will then occur so infrequently that they would not to be a problem. Improving the BER, however, adds significant additional data bits, thus increasing bandwidth and volume requirements, as well as requiring additional processing power. A related technique, however, to code different information channels with different degrees of error correction depending on their importance, has potential for increasing the effective BER without unduly increasing bandwidth, volume, or processing power requirements. Another technique to look for destruction of specific apriori known information about the data string due to error propagation in data compression also holds promise to allow detection and correction of errors missed through traditional error correction algorithms.

4. Recommendation

The most fruitful techniques will be ones where error containment and error correction are integrated with data compression to provide optimal performance for both. The error containment characteristics of existing compression schemes should be analyzed for their behavior under different data and error conditions. The error tolerance requirements of different data sets need to be understood, so guidelines can then be developed for matching error requirements to suitable compression algorithms. Work should be done to develop new compression algorithms, or modify existing compression algorithms, to improve error containment behavior. Work should also be done to look for ways in which data compression could aid error detection and subsequent error correction.

Participants

The participants in this discussion group were Mayun Chang, Kar-Ming Cheung, P. C. Hariharan, Ben Kobler, Joan S. Langdon, Edward Seiler, and Gregory S. Yovanof. See the appendix for addresses.

APPENDIX

omit to
end

WORKSHOP PARTICIPANTS

Space and Earth Science Data Compression Workshop
Snowbird, Utah - April 11, 1991

Held in conjunction with the Data Compression Conference (DCC'91)

David W. Abrahamson
Department of Computer Science
Trinity College
Dublin 2
Ireland
phone: +353-1-772941
fax:
email: david@cs.tcd.ie

Karen Anderson
IBM T. J. Watson Research Center
P. O. Box 704
Yorktown Heights, NY 10598
phone: (914) 784-6577
fax:
email:

Nuno Bandeira
Santa Barbara Research Center
75 Coromar Drive
MS B28/87
Goleta, CA 93117
phone:
fax:
email:

Joe Bredekamp
Code SM
NASA Headquarters
Washington, DC 20546
phone: (202) 453-1505
fax: (202) 755-9235
email: NASAMAIL - JBREDEKAMP
Internet - jbredekamp@nasamail.nasa.gov

Mayun Chang
Dept. of Computer Science
University of California Santa Cruz
Santa Cruz, CA 95064
or
513 Central Avenue, Apt. R
Mountain View, CA 94043
phone: (415) 694-7718
fax:
email: manyun@saturn.ucsc.edu

Kar-Ming Cheung
Jet Propulsion Laboratory
MS 238-420
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 393-9480
fax:
email: kmcheung@rtop71.jpl.nasa.gov

Wen-Kuang Chou
University of Hawaii
Holmes Hall No. 493
2540 Dale Street
Honolulu, HI 96826
phone: (808) 956-7249
fax:
email: wkchou@wiliki.eng.hawaii.edu

Pamela C. Cosman
Stanford University
Durand Bldg, 145B
Stanford, CA 94305-4055
phone: (415) 324-4259
fax:
email: pamela@isl.stanford.edu

PRECEDING PAGE BLANK NOT FILMED

David Craft
IBM Austin Development Lab
Dept F27S, Bldg 810, zip 3081
11400 Burnet Road
Austin, TX 78758
phone: (512) 838-9614
fax: (512) 838-9770
email:

Reiner Creutzburg
University of Karlsruhe
Institute of Algorithms and Cognitive Systems
Am Fasanengarten 5
P. O. Box 6980
D-7500 Karlsruhe
Federal Republic of Germany
phone: (+49)-721-608 5325
fax: (+49)-721-696 893

John C. Curlander
Jet Propulsion Laboratory
MS 300-235
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-8262
fax:
email:

Ahmed Desoky
EMACS
University of Louisville
Louisville, KY 40292
phone: (502) 588-6304
fax:
email: ahdeso01@ulkyvm

Greg Doherty
University of Wollongong
P. O. Box 1144
Wollongong 2500
Australia
phone: +61 42 213859
fax: +61 42 213262
email: greg@cs.uow.edu.au

Sam Dolinar
Jet Propulsion Laboratory
MS 238-420
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-7403
fax: (818) 354-6825
email:

Linda Jo Dolny
Optivision, Inc.
4009 Miranda Ave.
Palo Alto, CA 94304
phone: (415) 855-0200
fax: (415) 855-0222
email:

Jeff Dozier
NASA/Goddard Space Flight Center
Code 900
Greenbelt, MD 20771
phone: (301) 286-8228
fax: (301) 286-3884
email: dozier@crseo.ucsb.edu

Melanie Dutkiewicz
MacDonald Dettwiler Systems Division
13800 Commerce Parkway
Richmond, British Columbia
Canada V6V 2J3
phone: (604) 278-3411
fax:
email:

Michael Ehresman
Quay Communications
1700 Park Avenue, Suite 2500
P. O. Box 4229
Park City, UT 84060-4229
phone: (801) 486-3344
fax:
email: (mek@mda.ca ?)

Benjamin Epstein
David Sarnoff Research Center
CN5300
Princeton, NJ 08543-5300
phone: (609) 734-2073
fax: (609) 734-2044
email: bre@mrlsun.sarnoff.com

Daniel E. Erickson
Jet Propulsion Laboratory
MS 180-602
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-1656
fax: (818) 354-7354
email: derickson@nasamail [NASAMail]

Cole Erskine
Zoran Corporation
1705 Wyatt Drive
Santa Clara, CA 95054
phone: (408) 986-1314
fax:
email:

Wai-Chi Fang
Jet Propulsion Laboratory
MS
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-1648
fax:
email:

Robert G. Finch
South Dakota State University
Electrical Engineering Department
Box 2220
HH201
Brookings, SD 57007
phone: (605) 688-5217
fax:
email:

Jeffrey Freeman
TRW
92-3236
One Space Park
Redondo Beach, CA 90278
phone: (213) 813-0330
fax:
email:

Richard Frost
Brigham Young University
Dept. of Electrical & Computer Engineering
459 Clyde Building
Provo, UT 84602
phone: (801) 378-3930
fax:
email: rickf@ee.byu.edu

Daniel Glover
NASA/Lewis Research Center
MS 54-2
21000 Brookpark Road
Cleveland, OH 44135
phone: (216) 433-2847
fax: (216) 433-8705
email:

Robert Gray
Stanford University
Electrical Engineering Department
133 Durand Bldg
Stanford, CA 94305-4055
phone: (415) 723-4001, or -4539
fax: (415) 723-8473
email: gray@isl.stanford.edu

P. C. Hariharan
STX
7601 Ora Glen Drive, Suite 300
Greenbelt, MD 20770
phone: (301) 513-1650
fax: (301) 513-1608
email: hari@nssdca.gsfc.nasa.gov

William G. Hartz
Analex Corporation
NASA Lewis Research Center
3001 Aerospace Parkway
Brook Park, OH 44142
phone: (216) 977-0084
fax:
email:

Paul G. Howard
Brown University
Computer Science Department
Box 1910
Providence, RI 02912-1910
phone: (401) 863-7672
fax: (401) 863-7657
email: pgh@cs.brown.edu

A. Kris Huber
Utah State University
215 N. 100 East, No. 3
Logan, UT 84321
phone: (801) 752-0196
fax:
email: sl1z6@usu.edu

Lawrence Junker
NASA Kennedy Space Center, FL
CS-PSD-6
J. F. Kennedy Space Center, FL 32899
phone:
fax:
email:

Peter Kenny
NASA/Goddard Space Flight Center
Code 684.2
Greenbelt, MD 20771
phone: (301) 286-7445
fax:
email: kenny@uit.gsfc.nasa.gov

Rajiv Khanna
MITRE Corporation
7525 Colshire Drive
McLean, VA 22102-3481
phone: (703) 883-7758
fax:
email: rkhanna@mdf.mitre.ag

Dana Kloza
Southwest Research Institute
6220 Culebra
P. O. Drawer 28510
San Antonio, TX 78228-0510
phone: (512) 522-3989
fax:
email:

Ben Kobler
NASA/Goddard Space Flight Center
Code 935.1
Greenbelt, MD 20771
phone: (301) 286-3553
fax:
email: kobler@nssdc.gsfc.nasa.gov

Weidong Kou
NCR Canada Ltd.
580 Weber Street North
Waterloo, Ontario
Canada, N2J 4G5
phone: (519) 884-1710
fax: (519) 884-0610
email:

Joan S. Langdon
Bowie State University
Dept of Math and Computer Science
Bowie, MD 20721
phone: (301) 464-6662
fax:
email: langdonj@boe00.minc.umd.edu

Jun-ji Lee
M/S T1704
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-4993
fax: (818) 393-4089
email:

Eric Lin
Dept. of Computer Science
Brandeis University
Walthan, MA 02254
phone:
fax:
email:

Robert Lindsay
UNISYS
640 North 2200 West
Salt Lake City, UT 84116
phone: (801) 594-6626
fax:
email: bob@advtech.slc.unisys.com

Cheng-Chang Lu
Kent State University
Dept of Math and Computer Science
Kent, OH 44242
phone: (216) 673-1833
fax:
email: lucc@mcs.kent.edu

Manohar Mareboyana
NASA/Goddard Space Flight Center
Code 930.4
Greenbelt, MD 20771
phone: (301) 286-3397
fax: (301) 386-3221
email: manohar@chrpalg.gsfc.nasa.gov

V. John Mathews
Department of Electrical Engineering
University of Utah
Salt Lake City, UT 84112
phone: (801) 581-7869
fax: (801) 581-5281
email: mathews@ee.utah.edu

David McKibbin
Miciobose Software
180 Lakefront Drive
Hunt Valley, MD 21030
phone: (301) 771-6786
fax:
email:

Arshad Mian
GE Government Services
P.O. Box 138
Moffett Field, CA 94035
phone: (415) 604-3450
fax: (415) 604-6997
email: ammian@nasamail.nasa.gov

Kristo Miettinen
GE Astro-Space Division
GE Bldg
100 M2412
230 Goddard Blvd
King of Prussia, PA 19406
phone: (215) 354-5713
fax: (215) 354-2226
email:

Trent Mills
GE Houston NASA
1050 Bay Area Blvd
Houston, TX 77058
phone: (713) 483-5740
fax:
email: TMILLS [NASAMAIL]

Judy Mosley
Hughes-Santa Barabara Research Center
75 Coromar Drive
Goleta, CA 93117
phone: (805) 562-7578
fax:
email:

Amar Mukherjee
University of Central Florida
Department of Computer Science
Orlando, FL 32816-0362
phone: (407) 823-2763
fax:
email: amer@cs.ucf.edu

Jeff Niehaus
Texas Instruments
4032 Kentshire Lane
Dallas, TX 75287
phone: (214) 917-7954
fax:
email: niehaus@vdl.ti.com

Dmitry A. Novik
[NASA GSFC]
6225 Springhill Court, #304
Greenbelt, MD 20770
phone: (301) 345-2464
fax:
email: novik@chrpisis.gsfc.nasa.gov

Karen Oehler
Stanford University
Information Systems Laboratory
145A Durand Bldg
Stanford, CA 94305-4055
phone: (415) 723-2675
fax:
email: oehler@isl.stanford.edu

Ivan Onyszchuk
Jet Propulsion Laboratory
MS 238-420
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-3674
fax:
email: ivan@rtop71.jpl.nasa.gov

David C. Page
Computer Science Corporation
MS 242
3160 Fairview Park Drive
Falls Church, VA 22042
phone: (703) 876-1318
fax:
email:

Christopher J. Pestak
Analex Corporation @
NASA Lewis Research Center
3001 Aerospace Parkway
Brook Park, OH 44142
phone: (216) 977-0093
fax:
email:

Gary Promhouse
IBM
844 Don Mills Road
Toronto, Ontario
Canada
phone: (416) 448-3587
fax:
email: prom@csd.uwo.ca

Lew Randerson
Princeton Plasma Physics Lab
P. O. Box 451
Princeton, NJ 08543
phone: (609) 243-3134
fax:
email: randerson@usc.pppl.gov

N. Ranganathan
University of South Florida
Dept of Computer Science and Engineering
ENG 118
Tampa, FL 33620
phone: (813) 974-4760
fax:
email: nanganat@sol.usf.edu

Dale Rickman
Hughs Aircraft Company
Suite 100
7375 Executive Place
Seabrook, MD 20706
phone: (301) 805-0329
fax: (301) 805-0327
email: none

Eve Riskin
Department of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195
phone: (206) 685-2313
fax:
email: riskin@isdl.ee.washington.edu

Vincent Salomonson
NASA/Goddard Space Flight Center
Code 900
Greenbelt, MD 20771
phone: (301) 286-8601
fax: (301) 286-3884
email: VSALOMONSON [GSFCMAIL]

Rick Schumeyer
Naval Research Lab.
Code 4104
4155 Overlook Ave.
Washington, DC
phone: (202) 404-7966
fax:
email: ricks@luke.nrl.navy.mil

Edward Seiler
[ST Systems Corporation]
NASA/Goddard Space Flight Center
Code 930.4
Greenbelt, MD 20771
phone: (301) 794-5405
fax:
email: seiler@amarna.gsfc.nasa.gov

Sylvia Shen
Lockheed Palo Alto Research Lab.
O/96-40, B/2546
3251 Hanover Street
Palo Alto, CA 94304
phone: (415) 354-5019
fax:
email: dipl!shen@apple.com

S. Srikanth
R&D, CMC LTD.
115, S D Road
Secunderabad -500 003
INDIA
phone: (0842) 844476
fax:
email:

Robert Stack
GE Astro-Space Division
GE Bldg
100 M2450
230 Goddard Blvd
King of Prussia, PA 19406
phone: (215) 354-2780
fax: (215) 354-2226
email:

Mary C. Stahl
Booz, Allen and Hamilton
1760 Business Center Drive
Reston, VA 22090
phone: (703) 438-5068
fax:
email:

Virginia Stonick
Carnegie Mellon University
Department of Electrical/Computer Engineering
Pittsburgh, PA 15213
phone: (412) 268-6636
fax: (412) 268-3890
email: ginny@ece.cmu.edu

Wallace S. Tai
Jet Propulsion Laboratory
MS 301-235
4800 Oak Grove Drive
Pasadena, CA 91109
phone: (818) 354-7561
fax:
email:

James C. Tilton
NASA/Goddard Space Flight Center
Code 930.4
Greenbelt, MD 20771
phone: (301) 286-9510
fax: (301) 286-3221
email: tilton@chrpisis.gsfc.nasa.gov

Terry Torkelson
Eastern Washington University
Computer Science Department
MS 86
Cheney, WA 99004
phone: (509) 359-6016
fax:
email: ttork@ewu

Hsiao Sun
IBM/ISD
9211 Corporate Blvd
Rockville, MD 20880
phone: (301) 640-4653
fax:
email:

Val Vaughn
The Aerospace Corporation
P. O. Box 92957
Los Angeles, CA 90009
phone: (213) 336-1126
fax:
email: vaughn@aerospace.aero.org

Jack Venbrux
University of Idaho
NASA Space Engineering Research
Center for VLSI System Design
Moscow, ID 83843
phone: (208) 885-6023
fax: (208) 885-7579
email: jvenbrux@groucho.mrc.uidaho.edu

Jeff Vitter
Department of Computer Science
Brown University
Box 1910
Providence, RI 02912-1910
phone: (401) 863-7646
fax: (401) 863-7657
email: jsv@cs.brown.edu

Ray Walker
University of California
IGPP
Los Angeles, CA 90024-1567
phone: (213) 825-7685
fax:
email: rwalker@igpp.ucla.edu

Victor K. Wei
Bellcore
Room 2L-339
445 South Street
Morristown, NJ 07960
phone: (201) 829-4261
fax:
email: wekowe@bellcore.com

Wayne Whyte, Jr.
NASA/Lewis Research Center
21000 Brookpark Road
MS 54-2
Cleveland, OH 44135
phone: (216) 433-3482
fax: (216) 433-8705
email:

Ross N. Williams
Renaissance Software
16 Lerwick Avenue
Hazelwood Park 5066
South Australia, Australia
phone: 61 8 79-5020
fax:
email: ross@spam.ua.oz.au

Carol Wong
Lockheed Engineering and Sciences Co.
Engineering and Science Program
2400 NASA Road 1, C-22
P. O. Box 58561
Houston, TX 77258
phone: (713) 483-0056
fax:
email: wong@tcd.jsc.gov

James A. Woods
NASA/Ames Research Center [RIACS]
MS 230-5
Moffett Field, CA 94035
phone: (415) 762-7762
fax:
email: jam@riacs.edu

Gregory S. Yovanof
Kodak Berkeley Research
Eastman Kodak Company
2120 Haste Street
Berkeley, CA 94704
phone: (415) 649-2720
fax:
email:

Karen Zanter
EROS Data Center
Sioux Falls, SD 57198
phone: (605) 594-6848
fax:
email: zanter@pn9050.cr.usgs.gov

Marc Zipstein
C.E.R.I.L.
25 Cours B. Pascal
91000 Evry
FRANCE
phone:
fax:
email:

Report Documentation Page

NASA
National Aeronautics and Space Administration

1. Report No. NASA CP-3130	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Space and Earth Science Data Compression Workshop		5. Report Date November 1991
		6. Performing Organization Code 936
7. Author(s) James C. Tilton, Editor		8. Performing Organization Report No. 91B00149
		10. Work Unit No. 590-32-14-01
9. Performing Organization Name and Address Goddard Space Flight Center Greenbelt, Maryland 20771		11. Contract or Grant No.
		13. Type of Report and Period Covered Conference Publication
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546-0001		14. Sponsoring Agency Code

15. Supplementary Notes This workshop was organized and sponsored by the National Aeronautics and Space Administration (NASA) and cosponsored by the IEEE Computer Society Technical Committee on Computer Communications (TCCC). It was held in conjunction with the 1991 Data Compression Conference (DCC'91), which was sponsored by the IEEE TCCC in cooperation with NASA and the Center of Excellence in Space Data and Information Sciences at the NASA Goddard Space Flight Center.

16. Abstract This document is the proceedings from a "Space and Earth Science Data Compression Workshop," which was held on April 11, 1991, at the Snowbird Conference Center in Snowbird, Utah. This workshop was held in conjunction with the 1991 Data Compression Conference (DCC'91), which was held at the same location April 8-10, 1991. The workshop explored opportunities for data compression to enhance the collection and analysis of space and Earth science science data. The workshop focused on the scientists' data requirements, as well as constraints imposed by the data collection, transmission, distribution, and archival systems. The workshop consisted of several invited papers, followed by group discussions. Two invited papers described information systems for space and Earth science data. Four invited papers depicted analysis scenarios for extracting information of scientific interest from data collected by Earth-orbiting and deep-space platforms. A final invited paper was a general tutorial on image data compression. Most workshop participants joined one of three discussion groups covering (i) Data Compression for Data Archival and Browse/Quick Look, (ii) Data Compression for Near Earth and Deep Space to Earth Transmission, and (iii) Techniques for Containing Error Propagation in Compression/Decompression Schemes. The first goal of each discussion group was to examine the potential for data compression to address data storage and transmission constraints found throughout the domain of NASA missions. The second goal was to recommend specific actions directed at enabling mission use of appropriate data compression technologies to overcome these constraints. Each discussion group produced a report, which is included herein.

17. Key Words (Suggested by Author(s)) Data Compression, Image Compression Signal Processing, Image Processing Space Science, Earth Science		18. Distribution Statement Unclassified - Unlimited Subject Category 59	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 92	22. Price A05

END

DATE

FILMED

JAN 6 1992